Asymptotic Theory of Quantum Statistical Inference
Selected Papers

edited by
Masahito Hayashi

ERATO Quantum Computation and Information Project,
Japan Science and Technology Agency &
Graduate School of Engineering of Informatics,
University of Tokyo
PREFACE

In order to obtain information from a quantum system of interest, we need to perform a quantum measurement and extract the desired information from the obtained data. Needless to say, we had better optimize the above two processes for obtaining information of the quantum system. This research area is called “Quantum Statistical Inference,” and is much required for realizing many quantum information processing tasks. This research field was initiated in the middle of 1960s, and was studied up to 1980 by American and Russian researchers, for example Holevo, Yuen, Kennedy, and Belavkin, etc. Their research is summarized in the following books:

[a] C. W. Helstrom,
Quantum Detection and Estimation Theory,

[b] A. S. Holevo,
Probabilistic and Statistical Aspects of Quantum Theory,
(North-Holland, Amsterdam, 1982); Originally published in Russian (1980).

However, these researchers did not consider the asymptotic aspects while the asymptotic theory is essential for the large sample case in statistics. Moreover, an elegant general theory has been established in classical statistical inference theory.

In the middle of the 1980s, a different research direction has been started by Nagaoka (who is an expert in mathematical statistics and information geometry), which focused on the asymptotic theory. In the 1990s, several Japanese researchers (Fujiwara, Matsumoto, Ogawa, Hayashi) have been influenced by Nagaoka, and joined Quantum Statistical Inference. Hence, in the 1990s, by combining the mathematical formulation of quantum mechanics and mathematical statistics, these Japanese researchers obtained several good results in Quantum Statistical Inference. Especially, the Japanese
researchers have deeply discussed its asymptotic aspects, which had not been studied in the earlier stage. Recently, Quantum Statistical Inference has drawn the attention of several European statisticians (Gill, Bandorff-Nielsen, Jupp, Guta, Ballester, etc.) who joined this research field. On the other hand, several different directions of this research area were started in Europe after the 1990s by physicists, Massar, Popescu, D’Ariano, Bužek, Keyl, Werner, Bagan, Baig, Muñoz-Tapia, Gisin, Vidal, Latorre, Pascual, Tarrach etc. They were motivated by the foundations of physics.

However, despite this progress, there has been no English book covering the great progress in quantum statistical inference since Holevo’s textbook. Therefore, this book: *Asymptotic Theory of Quantum Statistical Inference* is expected to contribute not only to quantum information science and statistics but also to various other research fields.

Moreover, for the reader’s convenience, I add introductions (Chap 0 and introductions for every part), which explain preliminary concepts, historical backgrounds, and related researches. Indeed, some contributors strongly advised me to add introductions for the convenience of readers unfamiliar to quantum statistical inference, and agreed with these introductions. Particularly, a contributor strongly recommended me to introduce some researchers that are not contained in this book. However, these introductions are written based only on my personal point of view. Thus, I, the editor of this book, assume all responsibility for these parts. These introductions are different from the collective opinion of all contributors*. Therefore, they should be used only as reference materials for the historical significances of contained papers, and their significances should be decided by readers themselves.

I wish to thank Japan Science and Technology Agency for covering the reprint permission fee through ERATO Quantum Computation and Information Project. I would like to express my gratitude to Professor Hiroshi Imai and all members of Quantum Computation and Information Project for their warm support. I am particularly indebted to Professor Richard Gill, Professor Keiji Matsumoto, Professor Hiroshi Nagaoka, and Mr. Fuyuhiko Tanaka for a lot of helpful comments on the introductions. I am especially grateful to Mr. Fuyuhiko Tanaka, Dr. Kouki Yonezawa, Mr. Susumu Tokumoto, and Dr. Yosuke Kikuchi for supporting the editing process. Finally, my thanks go to all contributors and translators for their

*Indeed, a contributor sent me a strong objection against containing these introductions because he disagree with the content of these introductions.
Preface

contribution and careful proof reading.

Masahito Hayashi
ERATO, Quantum Computation and Information Project
Japan Science and Technology Agency
Hongo, Tokyo, Japan
viii

Asymptotic Theory of Quantum Statistical Inference
## CONTENTS

*Preface* ................................................................. v

First Appearance .......................................................... 1
List of Contributors ....................................................... 5
0 Introduction to Quantum Statistical Inference ...................... 8

**Part I: Hypothesis Testing** ............................................. 17

1 Introduction to Part I ................................................... 19
2 Strong Converse and Stein’s Lemma in Quantum Hypothesis Testing .................................................. 28
   *Tomohiro Ogawa and Hiroshi Nagaoka*
3 The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability ......................... 43
   *Fumio Hiai and Dénes Petz*
4 Strong Converse Theorems in Quantum Information Theory .................................................. 64
   *Hiroshi Nagaoka*
5 Asymptotics of Quantum Relative Entropy from a Representation Theoretical Viewpoint ................. 66
   *Masahito Hayashi*
6 Quantum Birthday Problems: Geometrical Aspects of Quantum Random Coding .................................. 75
   *Akio Fujiwara*

**Part II: Quantum Cramér-Rao Bound in Mixed States Model** ................................................. 89

7 Introduction to Part II .................................................. 91
Asymptotic Theory of Quantum Statistical Inference

8 A New Approach to Cramér-Rao Bounds for Quantum State Estimation .......................... 100 Hiroshi Nagaoka

9 On Fisher Information of Quantum Statistical Models ........................................ 113 Hiroshi Nagaoka

10 On the Parameter Estimation Problem for Quantum Statistical Models ........................ 125 Hiroshi Nagaoka

11 A Generalization of the Simultaneous Diagonalization of Hermitian Matrices and its Relation to Quantum Estimation Theory .................................................. 133 Hiroshi Nagaoka

12 A Linear Programming Approach to Attainable Cramér-Rao Type Bounds ................. 150Masahito Hayashi

13 Statistical Model with Measurement Degree of Freedom and Quantum Physics .................. 162Masahito Hayashi and Keiji Matsumoto

14 Asymptotic Quantum Theory for the Thermal States Family ........................................ 170Masahito Hayashi

15 State Estimation for Large Ensembles ........................................................................ 178Richard D. Gill and Serge Massar

Part III: Quantum Cramér-Rao Bound in Pure States Model 215

16 Introduction to Part III ......................................................................................... 217

17 Quantum Fisher Metric and Estimation for Pure State Models ............................ 220 Akio Fujiwara and Hiroshi Nagaoka

18 Geometry of Quantum Estimation Theory .......................................................... 229 Akio Fujiwara

19 An Estimation Theoretical Characterization of Coherent States .......................... 287 Akio Fujiwara and Hiroshi Nagaoka

20 A Geometrical Approach to Quantum Estimation Theory ...................................... 305 Keiji Matsumoto
Part IV: Group Symmetric Approach to Pure States Model

21 Introduction to Part IV ...................................... 353

22 Optimal Extraction of Information from Finite Quantum Ensembles ........................................... 356
Serge Massar and Sandu Popescu

23 Asymptotic Estimation Theory for a Finite-Dimensional Pure State Model ..................................... 365
Masahito Hayashi

24 Optimal Universal Quantum Cloning and State Estimation ......................................................... 379
Dagmar Bruß, Artur Ekert, and Chiara Macchiavello

25 Bounds for Generalized Uncertainty of Shift Parameter ............................................................... 386
Alexander S. Holevo

Part V: Large Deviation Theory in Quantum Estimation 393

26 Introduction to Part V .............................................. 395

27 On the Relation between Kullback Divergence and Fisher Information: From Classical Systems to Quantum Systems ............................................................. 399
Hiroshi Nagaoka

28 Two Quantum Analogues of Fisher Information from a Large Deviation Viewpoint of Quantum Estimation ................................................................. 420
Masahito Hayashi

29 Estimating the Spectrum of a Density Operator ................................................................. 458
Michael Keyl and Reinhard F. Werner

Part VI: Further Topics on Quantum Statistical Inference 469

30 Introduction to Part VI .............................................. 471

31 Optimal Quantum Clocks ........................................ 477
Vladimír Bužek, R. Derka, and Serge Massar

32 Quantum Channel Identification Problem ......................................................... 487
Akio Fujiwara

33 Homodyning as Universal Detection ................................................................. 494
Giacomo Mauro D’Ariano
34 On the Measurement of Qubits.......................... 509

Daniel F. V. James, Paul G. Kwiat, William J. Munro, and
Andrew G. White

Index......................................................... 539
FIRST APPEARANCE

Part I


Part II


Part III


Part IV


Part V

Chap. 27: H. Nagaoka, “On the relation between Kullback divergence and Fisher information — from classical systems to quantum


**Part VI**


LIST OF CONTRIBUTORS

Dagmar Bruß  
*Chap. 24*  
Inst. f. Theoret. Physik  
Universität Hannover  
Appelstr. 2, D-30167  
Hannover, Germany  
bruss@itp.uni-hannover.de

Vladimír Bužek  
*Chap. 31*  
Institute of Physics  
Slovak Academy of Sciences  
Dubravska cesta 9, 842 28  
Bratislava, Slovakia  
buzek@savba.sk

Giacomo Mauro D’Ariano  
*Chap. 33*  
Dipartimento di Fisica “A. Volta”  
via Bassi 6, I-27100 Pavia, Italy  
dariano@unipv.it

R. Derka  
*Chap. 31*  

Artur Ekert  
*Chap. 24*  
Centre for Quantum Computation  
University of Cambridge  
Cambridge CB3 0WA, UK  
artur.ekert@qubit.org

Akio Fujiwara  
*Chap. 6,17,18,19,32*  
Department of Mathematics  
Graduate School of Science  
Osaka University  
Toyonaka, Osaka 560-0043, Japan  
fujiwara@math.wani.osaka-u.ac.jp

Richard Gill  
*Chap. 15*  
Mathematical Institute  
University Utrecht  
Box 80010, 3508 TA  
Utrecht, Netherlands  
richard.gill@math.uu.nl

Masahito Hayashi  
*Chap. 5,12,13,14,23,28*  
ERATO, Quantum Computation  
and Information Project, JST  
Bunkyo-ku, Tokyo 113-0033, Japan  
masahito@qci.jst.go.jp

Fumio Hiai  
*Chap. 3*  
Division of Mathematics  
Graduate School of  
Information Sciences  
Tohoku University  
Aoba-ku, Sendai 980-8579, Japan  
hiai@math.is.tohoku.ac.jp
List of Contributors

Alexander S. Holevo
Chap. 25
Steklov Mathematical Institute
Gubkina 8, 119991 Moscow, Russia
holevo@genesis.mi.ras.ru

Daniel F. V. James
Chap. 34
Theoretical Division T-4 MS B-283
Los Alamos National Laboratory
Los Alamos, NM 87545, USA
dfvj@t4.lanl.gov

Michael Keyl
Chap. 29
Institut fur Mathematische Physik
TU-Braunschweig
Mendelssohnstrasse 3
38106 Braunschweig, Germany
m.keyl@tu-bs.de

Paul G. Kwiat
Chap. 34
University of Illinois
1110 W. Green St., Urbana
IL 61801-3080, USA
kwiat@uiuc.edu

Chiara Macchiavello
Chap. 24
Dipartimento di Fisica “A.Volta”
Università di Pavia
27100 Pavia, Italy

Serge Massar
Chap. 15,22,31
Service de Physique Théorique
CP 225
Université Libre de Bruxelles
Bvd. du Triomphe, B-1050
Bruxelles, Belgium
smassar@ulb.ac.be

William J. Munro
Chap. 34
Hewlett-Packard Laboratories
Filton Road, Stoke Gifford
Bristol BS34 8QZ, UK
bilmun@hplb.hpl.hp.com

Hiroshi Nagaoka
Chap. 2,4,8,9,10,11,17,19,27
Graduate School of
Information Systems
The University of
Electro-communications
Chofu-shi, Tokyo 182-8585, Japan
nagaoka@is.uec.ac.jp

Tomohiro Ogawa
Chap. 2
Dept. of Mathematical Informatics
Graduate School of Information
Science and Technology
University of Tokyo
Bunkyo-ku, Tokyo 113-8656, Japan
ogawa@sr3.t.u-tokyo.ac.jp
Dénés Petz  
Chap. 3  
Department for Mathematical Analysis  
Mathematical Institute  
Budapest University of Technology and Economics  
Megyetem rakpart 3-9, H-1111  
Budapest, Hungary  
petz@math.bme.hu

Sandu Popescu  
Chap. 22  
H H Wills Physics Lab.  
Bristol University  
Tyndall Av., Bristol BS8 1TL, UK  
s.popescu@bristol.ac.uk

Andrew G. White  
Chap. 34  
Department of Physics  
University of Queensland  
Brisbane, QLD 4072, Australia  
andrew.white@uq.edu.au

Reinhard F. Werner  
Chap. 29  
TU Braunschweig  
Institut fur Mathematische Physik  
Mendelsohnstr. 3  
38106 Braunschweig, Germany  
r.werner@tu-bs.de

Translators of Nagaoka’s paper:  

Fuyuhiko Tanaka  
Chap. 9,11  
Dept. of Mathematical Informatics  
Graduate School of Information Science and Technology  
University of Tokyo  
Bunkyo-ku, Tokyo 113-8656, Japan  
ftanaka@stat.t.u-tokyo.ac.jp

Yoshiyuki Tsuda  
Chap. 9,11  
21COE, Chuo University  
Bunkyo-ku, Tokyo 112-8551, Japan  
tsuda@tamacc.chuo-u.ac.jp
CHAPTER 0
Introduction to Quantum Statistical Inference

1. Classical/Quantum Statistical Inference

Quantum mechanics has a statistical aspect at its core. Suppose that the state $\rho$ of the given particle is unknown to us. If we want to fully know the density matrix of $\rho$, we need an infinite ensemble of quantum particles prepared in the same state. However, infinite ensembles do not exist in practice. Given a finite ensemble of identically prepared particles, how well can we estimate the state from the finite ensemble using results of quantum measurement? This problem is called quantum statistical inference, and the same problem for probability distributions has already been studied in mathematical statistics. Researchers in this field have investigated the problem of determining the unknown probability distribution from given finite data coming from it. (Statistical inference for probability distributions is called “classical” statistical inference in contrast with “quantum” statistical inference.) Thus, quantum statistical inference can be regarded as the quantum extension of classical statistical inference. Therefore, in order to comprehend it, we should first briefly outline the classical statistical inference theory.

1.1. Buildup of Classical Statistical Inference

Usually researchers in classical statistical inference divide their problems into two classes: estimation and hypothesis testing. In estimation, they estimate the unknown distribution from given finite data. Usually, they assume that the unknown probability distribution belongs to a certain parameterized subset, i.e., a family of probability distributions. Using this assumption, they replace their problem by the estimation of the parameter identifying the distribution. On the other hand, in hypothesis testing, they test whether the given hypothesis is true or not.

The formulations of classical estimation are given as follows: In classical statistical inference, we often assume that the data are independent,
identically distributed (i.i.d.) with the unknown probability distribution $p^\ast$. Based on this assumption, we can easily treat estimation theory in the case where the number $n$ of data is large, \textit{i.e.}, in the large sample case. This is because asymptotic expansions are available in the large sample case.

Usually, we evaluate the estimation error by the mean square error (MSE). The MSE of an estimator is almost proportional to $1/n$ in the large sample case. Hence, in this case, it is suitable to focus on the coefficient of the term $1/n$. The optimization theory of this coefficient is called "asymptotic optimality theory," (or "asymptotic theory") and is one of the main topics in classical estimation theory. Fortunately, a unified treatment of asymptotic theory is available, which is a great advantage.

On the other hand, there is no unified approach for the small sample case, \textit{i.e.}, the case where the number $n$ of data is small. Hence, several methods are proposed as follows. One is the \textit{Bayesian approach}, in which it is assumed in addition that the parameter is a random variable with a known distribution. This prior distribution is decided based on prior knowledge\footnote{The optimal performance in the Bayesian approach almost coincides with those in non-Bayesian approaches in the large sample case.}. However, we cannot use the Bayesian approach when our prior knowledge is insufficient to decide the prior distribution.

On the other hand, there are several non-Bayesian approaches, \textit{e.g.}, \textit{minimum variance unbiased estimation}, \textit{minimax} method, \textit{invariance approach}, and so on. In the minimum variance unbiased estimation, one demands unbiasedness: the expectation of the estimator always equals the true parameter; and one minimizes the error of the estimators satisfying this condition. However, there is no good practical reason to restrict our estimators to unbiased estimators. In the minimax method, one focuses on the worst error for an estimator given the family of probability distributions. This method provides an attractive estimator when the family of probability distribution has a homogenous structure. Otherwise, unfortunately, the optimal solution is not necessarily an attractive estimator. Further, the invariance approach is useful for a homogenous family of probability distribution, but is not so useful in other families. Thus, there is no one method that can always be successfully applied in the small sample case, \textit{i.e.}, there is no unified approach \cite{0-3].

Similarly, we can consider these methods for hypothesis testing, but

\footnote{Currently, many mathematical statisticians treat dependent data, and they obtained several basic results similar to the i.i.d. case. However, the discussion of the dependent case is so difficult that we mainly focus on the i.i.d. case in this book.}
they suffer the same drawbacks.

1.2. Research Direction of Quantum Statistical Inference

Due to the success of asymptotic theory in classical statistical inference, it is natural to attempt to establish an asymptotic theory in quantum statistical inference. Since we can easily apply classical asymptotic theory to the i.i.d. case, we need the following repeatability in quantum statistical inference:

- The unknown state can be prepared independently and identically.

This condition is a quantum analogue of the i.i.d. condition and is called quantum i.i.d. condition. It will be more formally formulated in the final section in this chapter.

Recently, these desired asymptotic theories have been established by reducing our problem in the large sample case as follows. First, the process of estimating/testing the unknown quantum state is divided into two parts: One is the appropriate choice of the quantum measurement, the other is that of the function providing the desired decision from the obtained data. While the latter belongs to the problem of classical statistics, the former is the central issue in quantum statistical inference. When we perform the same quantum measurement on every system, we can apply classical asymptotic theory. Therefore, our problem is reduced to the optimization of a suitable information quantity determined by our measurement‡.

2. Significance of Classical/Quantum Statistical Inference

We now consider what circumstances enable quantum statistical inference to be useful. First, we should remark that the importance of classical statistical inference depends on the amount of data. For example, if the number of observations is sufficiently huge (e.g., greater than 1,000,000,000) in a simple system, the obtained histogram almost equals the true distribution so that more technical analysis is not required§. On the other hand, if the number of data is not so huge (e.g., 1,000), by choosing a suitable estimator, we can improve the error. Especially, if the number of data is not

‡Even in the quantum i.i.d. case, when we use a correlated measurement on the composite system, we cannot use this reduction.
§If the model is complex and the amount of data is huge, we need theoretical analysis for treating data.
so huge but sufficiently large, we can apply asymptotic theory so that the approximately optimal error can be simply characterized\(^1\).

On the quantum side, it is assumed that the ensemble of quantum systems is prepared independently and identically. Quantum statistical inference is powerful when the size of the ensemble is not huge.

Such a situation often arises in quantum information processing. That is, a quantum state generator is needed in quantum information processing, and quantum statistical inference is a very useful method for identifying the quantum state \(\rho\) generated by it. Indeed, this method is more useful in the remote preparation of an entangled state because an identical state can be repeatedly produced; but the number of repetitions is limited.

Moreover, we can also consider applications to experiments with elementary particles. In this research area, it is often important to precisely estimate physical constants through experiments with energy restrictions\(^2\). In these experiments, only a limited amount of data is available so that quantum statistical inference can be expected to provide a guideline for the experiment.

However, quantum statistical inference (especially, its asymptotic theory) is not necessarily effective for every case. For example, the usual asymptotic theory of quantum statistical inference is not useful in the detection process in optical communication, although quantum statistical inference was initiated in relation to its optimization. This is because the quantum analogue of the i.i.d. setting is not satisfied in this case. But, as is mentioned in Sec. 1.3 of Chap. 1, a modified application is appropriate for the detection process [I-7,I-8].

As another example, we can cite the NMR system. In this case, the ensemble is so huge that we can treat not each individual particles but only the ensemble of particles, \(i.e.,\) we obtain only the average value of a measured observable with the ensemble as the data. Moreover, the huge size of the ensemble guarantees that the obtained average value almost equals the true average. Therefore, quantum statistical inference has little use in this case.

\(^1\)The complexity of a suitable estimator in the sense of mathematical statistics is also huge in the huge sample case. But, the complexity is not so huge when the number of data is not so huge but sufficiently large.

3. Organization of This Book

This book consists of the following parts

**Part I** Hypothesis Testing
**Part II** Quantum Cramér-Rao Bound I (Mixed States Model)
**Part III** Quantum Cramér-Rao Bound II (Pure States Model and Geometrical Study)
**Part IV** Group Symmetric Approach to Pure States Model
**Part V** Large Deviation Theory in Quantum Estimation
**Part VI** Further Topics on Quantum Statistical Inference

Every part begins with an introduction treating the relations between the included papers and related research. As is mentioned in Preface, these introductions are written based on the editor’s personal point of view, and explain related papers.

Part I consists of papers concerning quantum hypothesis testing, especially on quantum Stein’s lemma, and the remaining parts consist of papers concerning quantum estimation.

Parts II and III discuss the asymptotic theory of quantum estimation. Part II treats its general theory and the mixed state case. Part III treats the pure state case. In particular, Part III discusses the geometrical approach, which is effective for the pure state case and even for the mixed state case using the method of unbiased estimators.

Part IV treats the group symmetry approach in the pure state case, in which the minimax method is effective even in the small sample case. Its optimal solution is also the optimal Bayesian estimator under the invariant distribution. This part also discussed the relation between the small sample case and the large sample case.

Part V discusses large deviation theory of quantum estimation. This topic is closely related to hypothesis testing discussed in Part I.

Part VI treats the remaining two topics, the estimation of a quantum process and estimation by a fixed quantum measurement. The latter is often called tomography. Chapter 33 treats the estimation of an unknown state with the tomographic method in an infinite-dimensional system. Chapter 34 considers the application of quantum estimation to real physical systems.

The reader interested in the mathematical aspects of quantum statistical inference may read Chapter 1 first. The reader interested in the practical aspects of quantum statistical inference may read Chapter 34 first. This chapter explains its experimental significance.
Many papers in quantum statistical inference have appeared in the following e-print server**: http://arxiv.org/archive/quant-ph
whose preprints are abbreviated as quant-ph/???????

Several other preprints appeared in Mathematical Engineering Technical Reports, which is abbreviated as Math. Eng. Tech. Rep. These technical reports were previously managed by Mathematical Engineering Section of Department of Mathematical Engineering and Information Physics in School of Engineering, the University of Tokyo, and currently, by Department of Mathematical Informatics in Graduate School of Information Science and Technology, the University of Tokyo. Most of them are available at:

http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html

4. Mathematical Formulation

Here, we review the mathematical foundation of quantum measurement theory [0-1,0-2]. A quantum system is described by a Hilbert space $\mathcal{H}$, which is a finite or infinite dimensional linear space with an Hermitian inner product. Any state of the system $\mathcal{H}$ is described by a density matrix/operator $\rho$, which is a positive semi-definite Hermitian matrix/operator with trace 1.

The most general description of quantum measurement is provided by the mathematical concept of positive operator-valued measure (POVM) (or probability operator measure (POM)), which is a resolution of the unity on the system $\mathcal{H}$. Generally, the resolution $M = \{M(\omega)\}_{\omega \in \Omega}$ is called a POVM if it satisfies the following conditions††:

$$M(\omega) = M(\omega)^*, \quad M(\omega) \geq 0, \quad \sum_{\omega \in \Omega} M(\omega) = I.$$

When we perform the quantum measurement $M = \{M(\omega)\}_{\omega \in \Omega}$ on the system $\mathcal{H}$ whose state is $\rho$, the data $\omega$ obeys the probability distribution $P^M_\rho$:

$$P^M_\rho(\omega) = \text{Tr} \rho M(\omega).$$

The quantity $P^M_\rho(\omega)$ satisfies the axioms of probability as checked in the following case: The positivity of $\text{Tr} \rho M(\omega)$ follows from the positivity of $\rho$

**A contributor has an objection against introducing these preprints as further reading papers. However, the editor has decided to introduce them in order to provide the readers with more information.

††Some papers describe an outcome of a POVM by a subscript as $M = \{M_\omega\}_{\omega \in \Omega}$. 
and $M(\omega)$. The condition $\sum_{\omega \in \Omega} P^M_\rho(\omega) = 1$ is checked by the calculation: $\sum_{\omega \in \Omega} \text{Tr} \rho M(\omega) = \text{Tr} \rho \sum_{\omega \in \Omega} M(\omega) = \text{Tr} \rho I = \text{Tr} \rho = 1$. Thus, this description is sufficient to treat the output probability, but is not sufficient for the description of state demolition caused by the quantum measurement. Here, we do not treat the case when the set $\Omega$ is continuous, but its formulation in the continuous case is mentioned in section 2 of chapter 8.

Furthermore, when the two systems $\mathcal{H}_1$ and $\mathcal{H}_2$ are independently prepared in the states $\rho_1$ and $\rho_2$, respectively, the composite system is described by $\mathcal{H}_1 \otimes \mathcal{H}_2$, and its state is by $\rho_1 \otimes \rho_2$. Especially, when the $n$ systems are prepared independently and identically, the composite system is given by the $n$ tensor product space $\mathcal{H} \otimes \cdots \otimes \mathcal{H}$ and its state is described by an $n$-fold tensor product state $\rho \otimes \cdots \otimes \rho$. Since this can be regarded as a quantum analogue of the independent and identically distributed (i.i.d.) condition, it is called quantum i.i.d. condition. When we perform the same quantum measurement $M = \{M(\omega)\}_{\omega \in \Omega}$ on each of the components $\mathcal{H}$, the quantum measurement on the composite system $\mathcal{H} \otimes^{\otimes n}$ with $n$ outcomes $\omega_n \equiv (\omega_1, \ldots, \omega_n)$ is described by $M^n = \{M^n(\omega_n)\}$:

$$M^n(\omega_n) \equiv M(\omega_1) \otimes \cdots \otimes M(\omega_n).$$

We can also choose the $k$-th measurement $M_{k,\omega_{k-1}}$ as a function of the $k-1$ outcomes $\omega_{k-1} = (\omega_1, \ldots, \omega_{k-1})$ of the preceding measurements. This measurement on the composite system can be written by the POVM $M^n = \{M^n(\omega_n)\}$:

$$M^n(\omega_1, \ldots, \omega_n) \equiv \bigotimes_{k=1}^n M_{k,\omega_{k-1}}(\omega_k).$$

Such a POVM is called adaptive. On the other hand, if a POVM $M^n = \{M^n(\omega)\}$ has the following form, it is called separable:

$$M^n(\omega) = \sum_{x \in X_\omega} F_1(x) \otimes \cdots \otimes F_n(x).$$

Here, we should remark that $\{F_k(x)\}$ is not necessarily a POVM. Especially, a POVM $M^n = \{M^n(\omega)\}$ having the following form is separable.

$$M^n(\omega) = F_1(\omega) \otimes \cdots \otimes F_n(\omega).$$

Thus, any adaptive POVM is separable, i.e., the set of separable POVMs is a wider class of POVMs than that of adaptive POVMs. The set of separable POVMs also contains any POVM based on local operations and two-way
classical communications‡‡. Usually, it is very difficult to realize an arbitrary POVM on the composite system. Therefore, it is useful to find a good measurement among POVMs satisfying the same measurement condition (1) or the adaptive condition (2). As is mentioned in Sec. 4 of Chap. 7, the class (3) is important from a theoretical viewpoint.

For further introduction to quantum statistics, statistical researchers may read the survey paper [0-4], physical researchers may read Helstrom’s textbook [0-1], and mathematical researchers may read Holevo’s textbook [0-2]. Also, another book for a desired physical side of quantum statistical inference is planned [0-5].

References


PART I
Hypothesis Testing

Chap. 2: T. Ogawa and H. Nagaoka
“Strong converse and Stein’s lemma in quantum hypothesis testing” . . 28
Chap. 3: F. Hiai and D. Petz
“The proper formula for relative entropy and its asymptotics in
quantum probability” . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
Chap. 4: H. Nagaoka
“Strong converse theorems in quantum information theory” . . . . . . . 64
Chap. 5: M. Hayashi
“Asymptotics of quantum relative entropy from a representation
theoretical viewpoint” . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
Chap. 6: A. Fujiwara
“Quantum birthday problems: Geometrical aspects of quantum
random coding” . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75


CHAPTER 1

Introduction to Part I

1. Quantum Hypothesis Testing

1.1. Setting of Quantum Hypothesis Testing

In the quantum hypothesis testing, it is required to decide whether the given hypothesis (which is often called the null hypothesis) for the density is true or not. Especially, the following setting is the simplest, and is called simple hypothesis testing: the null hypothesis for the true density is given as a density $\rho$ and the alternative hypothesis (which is assumed to be valid if the null hypothesis is not valid) is also given as another density $\sigma$. Even in this simple setting, no quantum version of the fundamental asymptotic theory had been established until 1990. In the following sections, we briefly review the history of the development of its quantum version with an outline of its formulation. (For a more rigorous formulation, please refer to section 1 of [Chap. 2].)

In this problem, we focus on two error probabilities: the first error probability is the probability of incorrectly deciding $\sigma$ to be the true state when the true state is $\rho$, and the second error probability is the opposite error probability. In general, these two error probabilities do not have the same importance. Usually, we optimize the second error probability under the constraint that the first error probability is less than a given threshold.

In the classical version of this problem, i.e., the test of two probability distributions $p(\omega)$ and $q(\omega)$, the likelihood ratio $\frac{p(\omega)}{q(\omega)}$ plays an important role. In particular, as mentioned in Neyman Pearson Lemma, the optimal test is given by a likelihood ratio test, which is defined by the likelihood ratio.

In the quantum version, Holevo introduced a quantum counterpart of a likelihood ratio test, and obtained the quantum Neyman Pearson Lemma [I-1]. Following this result, Helstrom presented it in a more familiar form [I-2]. However, we cannot derive explicit forms of those error probabilities by only using this method.
1.2. Quantum Stein’s Lemma

In the classical asymptotic i.i.d. setting (i.e., with a sufficiently large i.i.d. classical ensemble), there is an explicit characterization for error probabilities as follows. Here, we focus on the constraint that the first error probability is lower than a given constant $\epsilon$. It is known that the optimal second error probability goes to 0 exponentially when the number of data goes to infinity. Furthermore, Stein’s Lemma guarantees that this exponential rate equals the relative entropy $^\ast D(p\|q) \equiv \sum_\omega p(\omega)(\log p(\omega) - \log q(\omega))$.

Using this lemma, we can characterize the asymptotic behavior of both of the two error probabilities. It should be noted that this decreasing rate is independent of the constant $\epsilon$.

As its quantum extension, one may think that this optimal decreasing rate is the quantum relative entropy $D(\rho\|\sigma) \equiv \text{Tr}\rho(\log \rho - \log \sigma)$ even in quantum hypothesis testing. But, the solution of this problem is not so easy. In the quantum setting, when we perform a quantum measurement $M$, the data obeys a probability distribution $P^M_\rho$. Thus, if we fix a quantum measurement $M$, our problem is reduced to the test of two probability distributions $P^M_\rho$ and $P^M_\sigma$. That is, if we perform the same measurement $M$ for the ensemble of the unknown state, the decreasing error rate equals classical relative entropy $D(P^M_\rho\|P^M_\sigma)$ by performing a suitable classical data manipulation. Here, we should remark that the monotonicity of the quantum relative entropy $^\dagger$ guarantees that $D(P^M_\rho\|P^M_\sigma) \leq D(\rho\|\sigma)$. Thus, for the optimal second error decreasing rate to be the quantum relative entropy, the achievement of the above inequality seems to be needed. But, in general, no quantum measurement $M$ is known to satisfy this equality. This fact seems to indicate that the optimal decreasing rate is less than the quantum relative entropy.

However, if it is proved that there exists a quantum measurement $M^n$ on an $n$-fold composite-system, such that the quantity $\frac{1}{n}D(P^{M^n}_{\rho\otimes n}\|P^{M^n}_{\sigma\otimes n})$ is sufficiently close to the quantum relative entropy $D(\rho\|\sigma)$, we can construct a test whose decreasing error rate is close to the quantum relative entropy $D(\rho\|\sigma)$. That is, the existence of such a test guaranteed by the application of the classical Stein’s Lemma to the hypotheses $P^{M^n}_{\rho\otimes n}$ vs. $P^{M^n}_{\sigma\otimes n}$ if there exists such a sequence of quantum measurements. Hiai and Petz [Chap.

---

*In statistics this is often called Kullback-Leibler divergence.

$^\dagger$Please refer to eq.(1) of Chap. 5.
proved its existence, i.e., by combining classical Stein’s lemma, they showed that the Quantum relative entropy can be achieved by the optimal second error exponent with the constant constraint. Undoubtedly, their achievement is a great progress in Quantum hypothesis testing.

However, Quantum Stein’s lemma was not completed only by the above result because an opposite inequality is required. In information theory, such a possibility part (achieving part, e.g., Hiai-Petz’s result) is called the direct part while the impossibility part is called the converse part. One may consider that the simple combination of the inequality $D(P^M_{\rho} \parallel P^M_{\sigma}) \leq D(\rho \parallel \sigma)$ and classical Stein’s Lemma would yield the converse part of quantum Stein’s Lemma. However, this is not true and more careful discussions are required, because in general, we must treat a sequence of correlated quantum measurements $M^n$, in which the probability distribution $P^{M^n}_{\rho \otimes n}$ is not an i.i.d. distribution.

Concerning the converse part of Quantum Stein’s lemma, Hiai and Petz derived its weak version, i.e., proved a quantum version of the weak converse, which will be explained in Sec. 3. After their success, Ogawa and Nagaoka succeeded in completely proving its converse part, i.e., the Quantum Stein’s Lemma has been proved [Chap. 2]. After this work, Nagaoka [Chap. 4] simplified its converse part.

### 1.3. Further Analysis of Quantum Stein’s Lemma

The first extension of Hiai-Petz’s result was performed by Hayashi [Chap. 5] with a representation-theoretical method. Their construction of $M^n$ satisfying $\frac{1}{n} D(P^{M^n}_{\rho \otimes n} \parallel P^{M^n}_{\sigma \otimes n}) \to D(\rho \parallel \sigma)$ depends on $\rho$ and $\sigma$. Hayashi showed that such a measurement $M^n$ can be chosen independently of $\rho$, and consequently obtained another proof of the direct part.

However, the above-mentioned proofs of quantum Stein’s Lemma require the classical Stein’s Lemma. Hence, these approaches did not directly construct the test achieving the optimal rate. Hayashi [I-3] found a pair of a measurement $M^n$ and a classical data process which achieves the optimal rate as follows: First, he characterized a measurement attaining this optimal rate as follows:

---

‡A trigger for the work of [Chap. 3] was the first meeting of Hiai and Nagaoka in 1990 when they were in the same university (but in different departments), in which Nagaoka asked the possibility of extending Stein’s lemma to the quantum case.

§Their method is quite simple, but their paper was written with the notations of operator algebra so that it may not be familiar to researchers in quantum information. This is because at that time there was few researchers in quantum information. Therefore, for the reader’s convenience, Ogawa-Nagaoka’s paper is treated as the first paper.
rate with a suitable likelihood test depending only on $\sigma$. He checked that the measurement proposed by Hayashi [Chap. 5] satisfies this condition. The main difficulty is that the probability $P_{\rho_n}^{M^n}$ is neither identical nor independent, and has been solved with the use of the information-spectrum method, which was introduced by Han\footnote{T. S. Han, “Hypothesis testing with the general source,” IEEE Trans. Inform. Theory, 46, no.7, 2415–2427 (2000).}. This method is a powerful tool for the analysis of classical hypothesis testing problems in such a general setting.

Nagaoka and Hayashi [I-5] focused on a quantum version of the information-spectrum approach of hypothesis testing, \textit{i.e.}, they asymptotically treated the simple quantum hypothesis testing problem with a general sequence of two density matrixes that are hypotheses, and obtained general formulas of this setting.

In the classical setting, Hoeffding\footnote{W. Hoeffding, “On probabilities of large deviations,” Proceedings of Symposium the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 203-219, Berkeley, University of California Press, 1965.} focused on the constraint on the decreasing rate of the second error probability, and obtained the optimal asymptotic decreasing rate of the first error probability under this constraint. Ogawa and Hayashi [I-4] proceeded to the analysis of its quantum version, and derived a lower bound of its optimal rate, which directly yields another proof of the direct part of quantum Stein’s Lemma without use of the classical Stein’s Lemma. This setting can be treated from the viewpoint of non-regular parametric estimation [I-6].

Furthermore, Ogawa and Nagaoka [I-7] remarked that the direct part of the channel coding theorem can be proved based on the direct part of Quantum Stein’s lemma. Following this research direction, Hayashi and Nagaoka [I-8] derived the capacity of the classical-quantum channel in a general setting. They also obtained a simple proof of Hiai-Petz’s result, at Remark 15, whose proof is based on more elementary knowledge.

1.4. \textbf{Further Topics in Quantum Hypothesis Testing}

Several quantum processings require the preparation of the maximally entangled state (Bell state). Thus, it is an interesting application of quantum hypothesis testing to test whether the experimentally realized state is truly the maximally entangled state or not. Several suitable measurements for this purpose have been proposed, and its performance and its optimality...
have been discussed by the papers [I-9,I-10,I-11].

On the other hand, van Dam et al. [I-12] discussed an optimal experiment for proving nonlocality of quantum mechanics. Furthermore, Osawa and Matsumoto [I-13] considered the time-energy uncertainty relation from the viewpoint of hypothesis testing.

2. Multi-Hypotheses Testing and Distinguishability

Since the number of assumed hypotheses is not necessarily two, we need to treat the multi-hypotheses case. Using a linear programming method, Yuen, Kennedy, and Lax [I-14] characterized the minimization problem of the average error probability with arbitrary prior probability. This result was generalized to convex optimization problems, e.g., the optimization of mutual information by Holevo [I-15].

When this average error is 0, the set of hypotheses states is said to be distinguishable. While the distinguishability of a set of pure states can be easily characterized by the orthogonality between the elements, its entangled version by local operation and classical communication (LOCC) is slightly complicated. Fan [I-16] obtained several remarkable results in this case. The asymptotic extension of distinguishability has a difficulty that is different from the entangled case. In order to analyze this problem, Fujiwara [Chap. 6] introduced two asymptotic extensions of orthogonality. The first extension is the asymptotic strong orthogonality, where the sum of all of the fidelities between distinct elements goes to 0. The second is the asymptotic weak orthogonality, where the value dividing the above sum by the number of elements goes to $0^\dagger$. When every element is randomly generated by a given probability, the maximum size satisfying the asymptotic weak orthogonality is characterized by the von Neumann entropy, but the maximum satisfying the asymptotic strong orthogonality is characterized by the quantum Rényi entropy with degree 2. Here, we should remark that as was solved by Hausladen et al. [I-17], the von Neumann entropy equals the capacity of quantum channel coding with pure states, and that the quantum Rényi entropy with degree 2 coincides with the smaller lower bound of the capacity by Stratonovich and Vantsjan [I-18]. Thus, Fujiwara’s results can

**Indeed, this result first appeared in H. P. Yuen, “Communication Theory of quantum systems,” Res. Lab. Electron., M.I.T., Cambridge Mass., Tech. Rep. 482, Aug. 1971. However, as Holevo pointed out to him, the first proof is invalid in the infinite-dimensional case. It was corrected in Yuen, Kennedy, and Lax [I-14].

††The definition of asymptotic weak orthogonality is seemingly different from the original one. But, they are essentially equivalent.
be regarded as their characterizations from the viewpoint of asymptotic weak/strong orthogonality.

3. Mathematical Treatment of Classical Hypothesis Testing

In classical simple hypothesis testing with probability distributions $p$ and $q$ on the probability space $\Omega$ under null and alternative, any test is described by the acceptance region $T$ for $p$, which is a subset of $\Omega$. Hence, the first error probability is $1 - p(T)$, and the second one is $q(T)$.

**Lemma 1: Neyman Pearson Lemma:** Assume that $1 - p(T_r) = \epsilon$. Then the equation

$$T_r = \arg\min_{1-p(T) \leq \epsilon} q(T)$$

(1)

holds, where the likelihood ratio test $T_r$ with the ratio $r$ is given by

$$T_r \equiv \{i \in \Omega | p_i \geq rq_i\}.$$  

Here, for the reader’s convenience, we give a proof of the classical Neyman-Pearson lemma.

**Proof:** Let $T$ be a test satisfying $1 - p(T) \leq \epsilon$. Thus,

$$\epsilon + rq(T_r) = 1 - p(T_r) + rq(T_r) = 1 + (rq - p)(T_r)$$

$$\leq 1 + (rq - p)(T) = 1 - p(T) + rq(T) \leq \epsilon + rq(T).$$

Therefore, we obtain (1).

Next, we proceed to the classical Stein’s lemma. Let $p^n$ and $q^n$ be the $n$-independent and identical distributions of $p$ and $q$, i.e., $p^n(\omega^n) \equiv p(\omega_1) \ldots p(\omega_n)$.

**Lemma 2: Stein’s Lemma:** The equation

$$\lim_{n} \frac{-1}{n} \log \beta_n^*(p||q|\epsilon) = D(p||q)$$

holds, where we define $\beta_n^*(p||q|\epsilon)$ as

$$\beta_n^*(p||q|\epsilon) \equiv \min_{1-p^n(T_n) \leq \epsilon} q^n(T_n).$$

††If we allow randomized likelihood test, the statement can be extended as follows: For a given $\epsilon > 0$, there exists a randomized likelihood test $T_r$ such that $1 - p(T_r) \leq \epsilon$ and $1 - p(T) \leq \epsilon \Rightarrow q(T_r) \leq q(T)$. In statistics, this extended statement is called Neyman Pearson Lemma.
 Indeed, classical Stein’s lemma can be divided into the following two parts: 

**Direct Part:** For any $\delta > 0$, there exists a sequence of acceptance regions $T_n$ of $n$ identical observations such that 

$$p^n(T_n) \to 0, \quad \lim \frac{-1}{n} \log q^n(T_n) \geq D(p\|q) - \delta.$$  

(2) 

**Strong) Converse Part:** 

$$\lim \frac{-1}{n} \log \beta_n^*(p\|q|\epsilon) \leq D(p\|q).$$  

(3) 

Combining the above statement, we can prove Stein’s lemma. The following statement is known as a weaker converse statement. 

**Weak Converse Part:** If $\epsilon_n \to 0$, then 

$$\lim \frac{-1}{n} \log \beta_n^*(p\|q|\epsilon_n) \leq D(p\|q).$$  

(4) 

Here, for reader’s convenience, we give a proof of classical Stein’s lemma. 

**Proof:** For arbitrary real number $\delta > 0$, we choose an acceptance region 

$$T_n = \{ \omega^n = (\omega_1, \ldots, \omega_n) \mid \frac{p_n}{q_n}(\omega^n) \geq e^{n(D(p\|q) - \delta)} \}.$$

Then, the first error probability 

$$1 - p^n(T_n) = p^n \left\{ \omega^n \left| \frac{1}{n} \sum_{i=1}^{n} (\log p - \log q)(\omega_i) < D(p\|q) - \delta \right. \right\}$$

goes to 0, because the random variable $\frac{1}{n} \sum_{i=1}^{n} (\log p - \log q)(\omega_i)$ goes to $D(p\|q)$ in $p^n$ probability. Thus, for a sufficiently large $N$, we have $1 - p^n(T_n) \leq \epsilon$, for any $n \geq N$. On the other hand, the second error probability satisfies 

$$q^n(T_n) = \sum_{\omega^n \in T_n} q^n(\omega_n) \leq \sum_{\omega^n \in T_n} e^{-n(D(p\|q) - \delta)} p^n(\omega_n)$$

$$\leq e^{-n(D(p\|q) - \delta)} \sum_{\omega^n} p^n(\omega_n) = e^{-n(D(p\|q) - \delta)}.$$

Thus, we obtain the direct part (2). 

Next, we show the opposite inequality 

$$\lim \frac{-1}{n} \log \beta_n^*(p\|q|\epsilon) \leq D(p\|q).$$  

(5)
Let $T_n$ be an acceptance region achieving the minimum value $\beta_n^*(p\|q|\epsilon)$, then

\[
\beta_n^*(p\|q|\epsilon) \geq q^n \left\{ \omega^n \in T_n \left| e^{n(D(p\|q)+\delta)} q^n(\omega^n) \geq p^n(\omega^n) \right. \right\} \\
\geq e^{-n(D(p\|q)+\delta)} p^n \left\{ \omega^n \in T_n \left| e^{n(D(p\|q)+\delta)} q^n(\omega^n) \geq p^n(\omega^n) \right. \right\} \\
\geq e^{-n(D(p\|q)+\delta)} \left( p^n(T_n) - p^n \left\{ \omega^n \left| e^{n(D(p\|q)+\delta)} q^n(\omega^n) \geq p^n(\omega^n) \right. \right\}^c \right),
\]

where $\delta$ is an arbitrary positive number. Since the random variable $\frac{1}{n} \sum_{i=1}^n (\log p - \log q)(\omega_i)$ goes to $D(p\|q)$ in the $p^n$ probability, the probability $p^n \left\{ \omega^n \left| e^{n(D(p\|q)+\delta)} q^n(\omega^n) \geq p^n(\omega^n) \right. \right\}^c$ goes to 0. This implies that $\left( p^n(T_n) - p^n \left\{ \omega^n \left| e^{n(D(p\|q)+\delta)} q^n(\omega^n) \geq p^n(\omega^n) \right. \right\}^c \right)$ goes to $1 - \epsilon > 0$. Therefore, we obtain

\[
\lim_{n \to \infty} -\frac{1}{n} \log \beta_n^*(p\|q|\epsilon) \leq D(p\|q) + \delta.
\]

Since $\delta$ is arbitrary, we obtain the converse part (3).

**Further Reading**


CHAPTER 2

Strong Converse and Stein’s Lemma
in Quantum Hypothesis Testing

Tomohiro Ogawa and Hiroshi Nagaoka

Abstract. The hypothesis testing problem for two quantum states is treated. We show a new inequality between the errors of the first kind and the second kind, which complements the result of Hiai and Petz to establish the quantum version of Stein’s lemma. The inequality is also used to show a bound on the probability of errors of the first kind when the power exponent for the probability of errors of the second kind exceeds the quantum relative entropy, which yields the strong converse in quantum hypothesis testing. Finally, we discuss the relation between the bound and the power exponent derived by Han and Kobayashi in classical hypothesis testing.

Index Terms: Quantum hypothesis testing, quantum information theory, quantum relative entropy, Stein’s lemma, strong converse.

1. Introduction

Let $\mathcal{H}$ be a Hilbert space which represents a physical system of interest. We assume $\text{dim } \mathcal{H} < \infty$ for mathematical simplicity. Let $\mathcal{L}(\mathcal{H})$ be the set of linear operators on $\mathcal{H}$ and denote the set of density operators on $\mathcal{H}$ by

$$
\mathcal{S}(\mathcal{H}) \overset{\text{def}}{=} \{ \rho \in \mathcal{L}(\mathcal{H}) \mid \rho = \rho^* \geq 0, \text{Tr}[\rho] = 1 \}.
$$

We study the hypothesis testing problem for the null hypothesis $H_0 : \rho \otimes^n \in \mathcal{S}(\mathcal{H} \otimes^n)$ versus the alternative hypothesis $H_1 : \sigma \otimes^n \in \mathcal{S}(\mathcal{H} \otimes^n)$, where $\rho \otimes^n$ and $\sigma \otimes^n$ are the $n$th-tensor powers of arbitrarily given density operators $\rho$ and $\sigma$ in $\mathcal{S}(\mathcal{H})$. The problem is to decide which hypothesis is true based on a two-valued quantum measurement $\{ A_n(H_0), A_n(H_1) \}$, where $A_n(H_0)$ and $A_n(H_1)$ are positive semidefinite operators on $\mathcal{H} \otimes^n$ obeying $A_n(H_0) + A_n(H_1) = I_n$ (the identity in $\mathcal{L}(\mathcal{H} \otimes^n)$). In the sequel, an operator $A_n \in \mathcal{L}(\mathcal{H} \otimes^n)$ satisfying $0 \leq A_n \leq I_n$, or a sequence $\{ A_n \}_{n=1}^{\infty}$ of such operators, is called a test, identifying it with the measurement $\{ A_n(H_0), A_n(H_1) \}$ =
\{A_n, I_n - A_n\}. For a test \(A_n\), the error probabilities of the first kind and the second kind are, respectively, defined by

\[
\alpha_n(A_n) \overset{\text{def}}{=} \text{Tr} [\rho^\otimes n (I_n - A_n)] \quad \text{and} \quad \beta_n(A_n) \overset{\text{def}}{=} \text{Tr} [\sigma^\otimes n A_n].
\]

We see that \(\alpha_n(A_n)\) is the probability of erroneously accepting \(\sigma^\otimes n\) when \(\rho^\otimes n\) is true and \(\beta_n(A_n)\) is the error probability of the converse situation. Our interest lies in the asymptotic behavior of these error probabilities when \(n\) goes to \(\infty\).

In [8], Hiai and Petz proved an important theorem which claims that the quantum relative entropy

\[
D(\rho||\sigma) \overset{\text{def}}{=} \text{Tr} [\rho (\log \rho - \log \sigma)]
\]

is obtained as a limit of classical relative entropies (Kullback divergences). Applying this theorem to the hypothesis testing problem, they showed that for any \(\varepsilon > 0\)

\[
\limsup_{n \to \infty} \frac{1}{n} \log \beta^*_n(\varepsilon) \leq -D(\rho||\sigma),
\]

where

\[
\beta^*_n(\varepsilon) \overset{\text{def}}{=} \min \{\beta_n(A_n) \mid A_n \in \mathcal{L}(\mathcal{H}^\otimes n), 0 \leq A_n \leq I, \alpha_n(A_n) \leq \varepsilon\}.
\]

In other words, for each \(\varepsilon > 0\) there exists a test \(\{A_n\}\) such that \(\alpha_n(A_n) \leq \varepsilon\) for all \(n\) and \(\beta_n(A_n) \sim e^{-nr}\) with \(r \geq D(\rho||\sigma)\) as \(n \to \infty\). Here we note that the \(\varepsilon\)-constraint on \(\alpha_n(A_n)\) can be strengthened to the convergence of \(\alpha_n(A_n)\) to 0 while keeping the rate condition on \(\beta_n(A_n)\) unchanged; i.e., there exists a test \(\{A_n\}\) satisfying

\[
\lim_{n \to \infty} \alpha_n(A_n) = 0 \quad \text{and} \quad \limsup_{n \to \infty} \frac{1}{n} \log \beta_n(A_n) \leq -D(\rho||\sigma).
\]

Indeed, it follows from (2) that for each \(\varepsilon > 0\) there exist a number \(n_0(\varepsilon)\) and a test \(\{A_n(\varepsilon) \in \mathcal{L}(\mathcal{H}^\otimes n)\}_{n=n_0(\varepsilon)}^{\infty}\) such that \(\alpha_n(A_n(\varepsilon)) \leq \varepsilon\) and \((1/n) \log \beta_n(A_n(\varepsilon)) \leq -D(\rho||\sigma) + \varepsilon\) for all \(n \geq n_0(\varepsilon)\). An example of \(\{A_n\}\) satisfying (4) is then given by

\[
A_n = A_n(\varepsilon_k) \quad \text{if} \quad n_0(\varepsilon_k) \leq n < n_0(\varepsilon_{k+1}),
\]

where \(\{\varepsilon_k\}_{k=1}^{\infty}\) is an arbitrary sequence of positive numbers converging monotonically to 0.

These results are categorized into the “direct part” in the sense that they are concerned with existence of good tests. For the “converse part”
concerned with nonexistence of too good tests, Hiai and Petz obtained

\[- \frac{1}{1 - \varepsilon} D(\rho \parallel \sigma) \leq \liminf_{n \to \infty} \frac{1}{n} \log \beta_n^*(\varepsilon). \tag{5}\]

They derived this from

\[nD(\rho \parallel \sigma) = D(\rho \otimes^n \parallel \sigma \otimes^n) \geq d(\alpha_n(A_n) \parallel 1 - \beta_n(A_n)) \]

\[= -h(\alpha_n(A_n)) - \alpha_n(A_n) \log(1 - \beta_n(A_n)) - (1 - \alpha_n(A_n)) \log \beta_n(A_n)\]

\[\geq - \log 2 - (1 - \alpha_n(A_n)) \log \beta_n(A_n), \tag{6}\]

where

\[d(p \parallel q) \overset{\text{def}}{=} p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q},\]

\[h(p) \overset{\text{def}}{=} -p \log p - (1 - p) \log(1 - p),\]

and the first inequality follows from the monotonicity of the quantum relative entropy [10, 15] (see also Remark 3 below). Note that there is a gap between (2) and (5).

In this paper* we will derive a new lower bound on the error probabilities which is tighter than (6), whereby the preceding results (2) and (5) will be complemented to establish the quantum version of Stein’s lemma (see e.g. [3], p. 115):

\[\lim_{n \to \infty} \frac{1}{n} \log \beta_n^*(\varepsilon) = -D(\rho \parallel \sigma).\]

Furthermore, the same bound will be applied to investigate the asymptotic behavior of \(\alpha_n(A_n)\) under the exponential-type constraint on \(\beta_n(A_n)\) and to prove the strong converse to (4) whose classical counterpart is found in [2, 7]; i.e., if \(\beta_n(A_n)\) goes to 0 with a rate \(\beta_n(A_n) \sim e^{-nr}\) for some \(r > D(\rho \parallel \sigma)\) then \(\alpha_n(A_n)\) necessarily goes to 1. Note that (6) can prove only the weak converse that \(\alpha_n(A_n)\) does not go to 0 under the same assumption. Finally, we will compare our result with that of Han and Kobayashi [7] in classical hypothesis testing.

2. A Bound on the Error Probabilities

In the sequel we assume that \(\text{Rng } \rho \subset \text{Rng } \sigma\), where \(\text{Rng}\) denotes the range of an operator, so that the relative entropy and the quantity \(\text{Tr}[\rho^{1+s} \sigma^{-s}]\) for \(s > 0\) are well defined. Obviously the assumption is satisfied if \(\rho\) and \(\sigma\) are strictly positive operators.

*A brief summary of the results was published in [12].
Let $\lambda$ be an arbitrary real number and let the spectral decomposition of the operator $\rho^{\otimes n} - e^{n \lambda} \sigma^{\otimes n}$ be written as
\[
\rho^{\otimes n} - e^{n \lambda} \sigma^{\otimes n} = \sum_{j \in J_n} \mu_{n,j} E_{n,j},
\]
where $\{\mu_{n,j}\}_{j \in J_n}$ are the eigenvalues and $\{E_{n,j}\}_{j \in J_n}$ are the operators representing the orthogonal projections onto the eigenspaces. Define a test $X_{n,\lambda}$ by
\[
X_{n,\lambda} \overset{\text{def}}{=} \sum_{j \in K_n} E_{n,j},
\]
where $K_n \overset{\text{def}}{=} \{j \mid \mu_{n,j} \geq 0\}$. Noting the equivalence
\[
\frac{1}{n} \log \frac{p^n(x^n)}{q^n(x^n)} \geq \lambda \iff p^n(x^n) - e^{n \lambda} q^n(x^n) \geq 0,
\]
we see that $X_{n,\lambda}$ is a quantum analog of the classical likelihood ratio test. Indeed, $X_{n,\lambda}$ turns out to be optimal in the sense of Neyman-Pearson or that of Bayes according to the quantum Neyman-Pearson lemma (see [6], p. 108) which is based on the following fact.

**Lemma 1:** Every test $A_n$ satisfies
\[
\text{Tr}[(\rho^{\otimes n} - e^{n \lambda} \sigma^{\otimes n})X_{n,\lambda}] \geq \text{Tr}[(\rho^{\otimes n} - e^{n \lambda} \sigma^{\otimes n})A_n].
\]

*Proof:*
\[
\text{Tr}[(\rho^{\otimes n} - e^{n \lambda} \sigma^{\otimes n})A_n] = \sum_{j \in J_n} \mu_{n,j} \text{Tr}[E_{n,j}A_n] \leq \sum_{j \in K_n} \mu_{n,j} \text{Tr}[E_{n,j}A_n]
\leq \sum_{j \in K_n} \mu_{n,j} \text{Tr}[E_{n,j}] = \text{Tr}[(\rho^{\otimes n} - e^{n \lambda} \sigma^{\otimes n})X_{n,\lambda}].
\]

We use this lemma to derive a new lower bound on the error probabilities of quantum tests.

**Theorem 2:** For any test $A_n$ and any $\lambda \in \mathbb{R}$ we have
\[
1 - \alpha_n(A_n) \leq e^{-n \varphi(\lambda)} + e^{n \lambda} \beta_n(A_n),
\]
where
\[
\varphi(\lambda) \overset{\text{def}}{=} \max_{0 \leq s \leq 1} \{\lambda s - \psi(s)\}
\]
\[
\psi(s) \overset{\text{def}}{=} \log \text{Tr}[\rho^{1+s}\sigma^{-s}].
\]
**Proof:** Define probability distributions \( p_n = \{p_{n,j}\}_{j \in J_n} \) and \( q_n = \{q_{n,j}\}_{j \in J_n} \) on the set \( J_n \) by

\[
p_{n,j} = \text{Tr}[\rho^\otimes n E_{n,j}] \quad q_{n,j} = \text{Tr}[\sigma^\otimes n E_{n,j}].
\]

Noting that \( \mu_{n,j} \text{Tr} [E_{n,j}] = p_{n,j} - e^{n\lambda} q_{n,j} \) follows from (7) and using the Markov inequality, we have for \( \forall s \geq 0 \)

\[
\text{Tr} [\rho^\otimes n X_{n,\lambda}] = \sum_{j \in K_n} p_{n,j} = \text{Pr} \left\{ \frac{p_{n,j}}{q_{n,j}} \geq e^{n\lambda} \mid J \sim p_n \right\}
\]

\[
\leq e^{-n\lambda s} \mathbb{E} \left[ \left( \frac{p_{n,j}}{q_{n,j}} \right)^s \mid J \sim p_n \right] = e^{-n\lambda s} \sum_{j \in J_n} p_{n,j}^{1+s} q_{n,j}^{-s}
\]

where \( J \sim p_n \) means that \( J \) is a random variable subject to \( p_n \). Now we restrict the range of \( s \) to \( 0 \leq s \leq 1 \) and invoke the following inequality (see Remark 3 below):

\[
\sum_{j \in J_n} p_{n,j}^{1+s} q_{n,j}^{-s} \leq \text{Tr} (\rho^\otimes n)^{1+s} (\sigma^\otimes n)^{-s}.
\]

We thus have

\[
\text{Tr} [\rho^\otimes n X_{n,\lambda}] \leq \min_{0 \leq s \leq 1} e^{-n\lambda s} \text{Tr} (\rho^\otimes n)^{1+s} (\sigma^\otimes n)^{-s} = \min_{0 \leq s \leq 1} e^{-n(\lambda s - \psi(s))} = e^{-n\varphi(\lambda)}.
\]

Combining this with (8), the theorem is proved as follows:

\[
1 - \alpha_n(A_n) = \text{Tr} [\rho^\otimes n A_n] \leq \text{Tr} [ (\rho^\otimes n - e^{n\lambda} \sigma^\otimes n) X_{n,\lambda} ] + e^{n\lambda} \text{Tr} [\sigma^\otimes n A_n] \leq \text{Tr} [\rho^\otimes n X_{n,\lambda}] + e^{n\lambda} \text{Tr} [\sigma^\otimes n A_n] \leq e^{-n\varphi(\lambda)} + e^{n\lambda} \beta_n(A_n).
\]

**Remark 3:** As a quantum extension of Csiszár’s \( f \)-divergence ([4]), Petz [13, 14] introduced a quantity, which we denote by \( S_f(\rho\|\sigma) \), for arbitrary density operators (or arbitrary states on an arbitrary von Neumann algebra) \( \rho, \sigma \) and an arbitrary operator convex function \( f : \mathbb{R}^+ \rightarrow \mathbb{R} \). Here we say that \( f \) is operator convex (see e.g. [1], p. 123) if

\[
f((1 - \lambda)A + \lambda B) \leq (1 - \lambda)f(A) + \lambda f(B)
\]

holds for all positive matrices \( A, B \), and all \( 0 \leq \lambda \leq 1 \). Using the spectral representations

\[
\rho = \sum_i a_i F_i \quad \sigma = \sum_j b_j G_j,
\]

(13)
where \( \{a_i\}, \{b_j\} \) are the eigenvalues and \( \{F_i\}, \{G_j\} \) are the orthogonal projections onto the eigenspaces, \( S_f(\rho \| \sigma) \) is represented as

\[
S_f(\rho \| \sigma) = \sum_{i,j} a_i f \left( \frac{b_j}{a_i} \right) \Tr[F_i G_j].
\] (14)

The definition of \( S_f(\rho \| \sigma) \) is summarized in Appendix A. The relative entropy \( D \) defined by (1) is an example of \( S_f \) corresponding to the operator convex function \( f(u) = -\log u \). Other important operator convex functions are \( f(u) = -u^t \) for \( 0 \leq t \leq 1 \) and \( f(u) = u^t \) for \( 1 \leq t \leq 2 \) or \(-1 \leq t \leq 0\), and the corresponding \( f \)-divergences are represented as

\[
S_f(\rho \| \sigma) = \begin{cases}
-\Tr[\rho^{1-t} \sigma^t], & \text{if } 0 \leq t \leq 1 \\
\Tr[\rho^{1-t} \sigma^t], & \text{if } 1 \leq t \leq 2 \text{ or } -1 \leq t \leq 0.
\end{cases}
\] (15)

Petz showed that the monotonicity property ([10, 15]) of the relative entropy can be extended to the class of \( S_f \). In particular, \( S_f \) is monotone with respect to an arbitrary measurement (positive operator-valued measure or POM) \( \{E_j\} \), which is a family of positive semidefinite operators satisfying \( \sum_j E_j = I \), in the following sense. Let \( \{p_j\} \) and \( \{q_j\} \) be the probability distributions determined from \( \rho, \sigma \) and \( \{E_j\} \) by \( p_j = \Tr[\rho E_j] \) and \( q_j = \Tr[\sigma E_j] \). Then the \( f \)-divergence

\[
S_f(p \| q) \overset{\text{def}}{=} \sum_j p_j f \left( \frac{q_j}{p_j} \right)
\]
is not greater than \( S_f(\rho \| \sigma) \). In the case of (15) we have

\[
\Tr[\rho^{1-t} \sigma^t] \leq \sum_j p_j^{1-t} q_j^t \quad \text{if } 0 \leq t \leq 1,
\] (16)

and

\[
\Tr[\rho^{1-t} \sigma^t] \geq \sum_j p_j^{1-t} q_j^t \quad \text{if } 1 \leq t \leq 2 \text{ or } -1 \leq t \leq 0.
\] (17)

Now we can see that inequality (12) is a special case of (17). We also note that the monotonicity of \(-\Tr[\rho^{1-t} \sigma^t]\) for \( 0 \leq t \leq 1 \) including (16) is closely related to the Wigner-Yanase-Dyson-Lieb concavity (e.g. [9]) and follows from the interpolation theory of Uhlmann ([15]) as well as from Petz’s theorem.

**Remark 4:** The proof of Theorem 2 is essentially based on an argument which is familiar in classical large deviation theory. It is interesting to compare it with the proof of Cramér’s theorem (e.g. [5], p. 26).
3. Properties of $\psi(s)$ and $\varphi(\lambda)$

Let us observe some properties of the functions $\psi(s)$ and $\varphi(\lambda)$ for the later argument. First, we have

$$
\frac{d}{ds} \left( e^{-\psi(s)} \rho^{1+s} \sigma^{-s} \right) = -e^{-\psi(s)} \psi'(s) \rho^{1+s} \sigma^{-s} + e^{-\psi(s)} \rho^{1+s} (\log \rho) \sigma^{-s} - e^{-\psi(s)} \rho^{1+s} (\log \sigma) \sigma^{-s} = e^{-\psi(s)} \rho^{1+s} A \sigma^{-s}
$$

where

$$
A \overset{\text{def}}{=} \log \rho - \log \sigma - \psi'(s).
$$

Since the trace of $e^{-\psi(s)} \rho^{1+s} \sigma^{-s}$ is identically equal to 1 by the definition of $\psi$, its derivative vanishes and we obtain from (18)

$$
\text{Tr}[\rho^{1+s} A \sigma^{-s}] = 0, \quad (19)
$$

or, equivalently,

$$
\psi'(s) = e^{-\psi(s)} \text{Tr}[\rho^{1+s} (\log \rho - \log \sigma) \sigma^{-s}] . \quad (20)
$$

This implies that $\psi'(0) = D(\rho||\sigma)$ since $\psi(0) = 0$. From (19) we also have

$$
0 = \frac{d}{ds} e^{-\psi(s)} \text{Tr}[\rho^{1+s} A \sigma^{-s}]
$$

$$
= \text{Tr} \left[ \left( \frac{d}{ds} e^{-\psi(s)} \sigma^{-s} \rho^{1+s} \right) A \right] + e^{-\psi(s)} \text{Tr} \left[ \rho^{1+s} \left( \frac{d}{ds} A \right) \sigma^{-s} \right].
$$

Substituting (18) and $(d/ds)A = -\psi''(s)$ into this, we have

$$
\psi''(s) = e^{-\psi(s)} \text{Tr}[\rho^{1+s} A \sigma^{-s} A]
$$

$$
= e^{-\psi(s)} \text{Tr} \left[ \left( \rho^{\frac{1+s}{2}} A \sigma^{-\frac{s}{2}} \right) \left( \rho^{\frac{1+s}{2}} A \sigma^{-\frac{s}{2}} \right)^* \right] \geq 0
$$

which means that $\psi(s)$ is a convex function.

Next, $\varphi(\lambda)$ is the Legendre transformation of $\psi(s)$ and hence is convex. Letting

$$
\hat{s}(\lambda) \overset{\text{def}}{=} \arg \max_{0 \leq s \leq 1} \{ \lambda s - \psi(s) \}
$$

so that $\varphi(\lambda) = \lambda \hat{s}(\lambda) - \psi(\hat{s}(\lambda))$, we can see from Figs. 2.1 and 2.2 the following facts.

(i) If $\lambda \leq \psi'(0) = D(\rho||\sigma)$,

$$
\hat{s}(\lambda) = 0 \quad \text{and} \quad \varphi(\lambda) = 0.
$$
(ii) If $D(\rho||\sigma) < \lambda < \psi'(1)$,

$$0 < \hat{s}(\lambda) < 1, \quad \varphi(\lambda) > 0 \quad \text{and} \quad \psi'(\hat{s}(\lambda)) = \lambda.$$ 

(iii) If $\lambda > \psi'(1)$,

$$\hat{s}(\lambda) = 1 \quad \text{and} \quad \varphi(\lambda) = \lambda - \psi(1) > 0.$$ 

Although not essential in the following, it may be worth pointing out that for $0 \leq s \leq 1$ we have the inverse Legendre transformation

$$\psi(s) = \max_{-\infty < \lambda < \infty} \{\lambda s - \varphi(\lambda)\} = \max_{D(\rho||\sigma) \leq \lambda \leq \psi'(1)} \{\lambda s - \varphi(\lambda)\}.$$
4. The Quantum Stein’s Lemma

**Theorem 5:** For any $0 < \varepsilon < 1$, it holds that

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_n^*(\varepsilon) = -D(\rho \| \sigma).$$  \hspace{1cm} (22)

**Proof:** Since we already have (2), it suffices to show that

$$\lim \inf_{n \to \infty} \frac{1}{n} \log \beta_n^*(\varepsilon) \geq -D(\rho \| \sigma).$$  \hspace{1cm} (23)

Let $A_n$ be an arbitrary test which satisfies $\alpha_n(A_n) \leq \varepsilon$. From (9), we have

$$1 - \varepsilon \leq 1 - \alpha_n(A_n) \leq e^{-n\varphi(\lambda)} + e^{n\lambda} \beta_n(A_n),$$

and hence

$$\beta_n(A_n) \geq e^{-n\lambda}(1 - \varepsilon - e^{-n\varphi(\lambda)}).$$

By taking the minimum with respect to $A_n$, we obtain

$$\beta_n^*(\varepsilon) \geq e^{-n\lambda}(1 - \varepsilon - e^{-n\varphi(\lambda)}).$$  \hspace{1cm} (24)

Assume now that $\lambda > D(\rho \| \sigma)$. Then $\varphi(\lambda) > 0$ and

$$1 - \varepsilon - e^{-n\varphi(\lambda)} > 0$$

for sufficiently large $n$. Thus (24) yields

$$\frac{1}{n} \log \beta_n^*(\varepsilon) \geq -\lambda + \frac{1}{n} \log(1 - \varepsilon - e^{-n\varphi(\lambda)}),$$

and hence

$$\lim \inf_{n \to \infty} \frac{1}{n} \log \beta_n^*(\varepsilon) \geq -\lambda.$$ 

Since $\lambda$ can be arbitrarily close to $D(\rho \| \sigma)$, (23) has been proved. \hfill \blacksquare

5. The Strong Converse Property

For each $r > 0$, there uniquely exists a number $\lambda^* = \lambda^*(r)$ such that $\varphi(\lambda^*) = r - \lambda^*$ (see Fig. 2.2), and we define

$$u(r) \overset{\text{def}}{=} \varphi(\lambda^*(r)) = r - \lambda^*(r).$$  \hspace{1cm} (25)

**Theorem 6:** For any test $\{A_n\}$ and any $r > 0$, if

$$\lim \sup_{n \to \infty} \frac{1}{n} \log \beta_n(A_n) \leq -r$$  \hspace{1cm} (26) \hfill \text{then}

$$\lim \sup_{n \to \infty} \frac{1}{n} \log(1 - \alpha_n(A_n)) \leq -u(r).$$  \hspace{1cm} (27)
Proof: Given an arbitrary $\delta > 0$, it follows from the assumption (26) that there exists a number $n_0$ such that
\[
\beta_n(A_n) \leq e^{-n(r - \delta)} \quad \forall n \geq n_0.
\]
Then letting $\lambda = \lambda^*(r)$ in (9), we have
\[
1 - \alpha_n(A_n) \leq e^{-n\varphi(\lambda^*)} + e^{-n(r - \delta - \lambda^*)} = e^{-nu(r)} + e^{-n(u(r) - \delta)}, \quad \forall n \geq n_0,
\]
and hence
\[
\limsup_{n \to \infty} \frac{1}{n} \log(1 - \alpha_n(A_n)) \leq -u(r) + \delta.
\]
Since $\delta > 0$ is arbitrary, (27) has been proved.

Now we can easily see that if $r > D(\rho || \sigma)$ then $u(r) > 0$ (see Fig. 2.2). Thus we conclude that the strong converse holds as follows.

**Corollary 7:** For any test $A_n$, if
\[
\limsup_{n \to \infty} \frac{1}{n} \log \beta_n(A_n) < -D(\rho || \sigma), \quad (28)
\]
then $\alpha_n(A_n)$ goes to 1 exponentially.

A compact expression of the function $u(r)$ is given in the following theorem. This is an analog of the representation of the rate function for classical hypothesis testing given by Blahut (Theorem 7 [3]), although what he treated there was not the exponent of $1 - \alpha_n(A_n)$ but that of $\alpha_n(A_n)$.

**Theorem 8:**
\[
u(r) = \max_{0 \leq s \leq 1} \left\{ \frac{s}{1 + s} r - \frac{1}{1 + s} \psi(s) \right\}.
\]  

Proof: For any $r > 0$ and any $0 \leq s \leq 1$, we obtain from (25)
\[
u(r) = \varphi(\lambda^*(r)) \geq \lambda^*(r)s - \psi(s) \quad (30)
\]
and
\[
u(r) = r - \lambda^*(r).
\]
Eliminating $\lambda^*(r)$ from these equations, we have
\[
u(r) \geq \frac{s}{1 + s} r - \frac{1}{1 + s} \psi(s). \quad (31)
\]
Now there exists a number $s$ in $0 \leq s \leq 1$ achieving the equality in (30) by the definition of $\varphi(\lambda)$, and the $s$ achieves the equality in (31) as well. Consequently, we obtain (29).
6. Relation with Classical Hypothesis Testing

In this section, we discuss the relation between \( u(r) \) and the power exponent derived by Han and Kobayashi [7] for classical hypothesis testing.

Let \( p \) and \( q \) be arbitrary probability distributions on a finite set \( \mathcal{X} \) and denote their \( n \)th independent and identically distributed (i.i.d.) extensions by \( p^n \) and \( q^n \). We consider the hypothesis testing problem for the hypotheses \( p^n \) and \( q^n \) and denote the error probabilities by \( \alpha_n(A_n) = p^n(A^n) \) and \( \beta_n(A_n) = q^n(A_n) \) for an acceptance region \( A_n \subset \mathcal{X}^n \). After Blahut [3] showed that if \( \beta_n(A_n) \leq e^{-nr} \) with \( r > D(p||q) \) then \( \alpha_n(A_n) \) necessarily tends to 1 as \( n \to \infty \) at least with a polynomial rate, Han and Kobayashi [7] proved that the actual rate is exponential and found the optimal exponent. Namely, letting

\[
\alpha_n^*(r) \overset{\text{def}}{=} \min \{ \alpha_n(A_n) \mid A_n \subset \mathcal{X}^n, \beta_n(A_n) \leq e^{-nr} \},
\]

they derived

\[
\lim \inf_{n \to \infty} \frac{1}{n} \log(1 - \alpha_n^*(r)) = -\tilde{u}(r), \tag{32}
\]

where

\[
\tilde{u}(r) \overset{\text{def}}{=} \min_{\hat{p} : D(\hat{p}||p) \leq r} \{ D(\hat{p}||p) + r - D(\hat{p}||q) \}. \tag{33}
\]

Now assume that \( r \) is greater than but sufficiently near \( D(p||q) \). Then as remarked in [7], the minimum of (33) is attained with \( D(\hat{p}||q) = r \) and hence we have

\[
\tilde{u}(r) = \min_{\hat{p} : D(\hat{p}||q) = r} D(\hat{p}||p). \tag{34}
\]

As shown in Appendix B, (34) is rewritten as

\[
\tilde{u}(r) = \max_{0 \leq s \leq 1} \left\{ \frac{s}{1 + s} r - \frac{1}{1 + s} \log \sum_{j \in \mathcal{X}} p_j^{1+s} q_j^{-s} \right\}, \tag{35}
\]

which just corresponds to (29).

It is interesting to observe that a direct translation of the argument of Appendix B for deriving (35) from (34) into the quantum setting yields

\[
\min_{\hat{\rho} : D(\hat{\rho}||\sigma) = r} D(\hat{\rho}||\rho) = \max_{0 \leq s \leq 1} \left\{ \frac{s}{1 + s} r - \frac{1}{1 + s} \overline{\psi}(s) \right\}
\]

where

\[
\overline{\psi}(s) \overset{\text{def}}{=} \log \text{Tr} \left[ e^{(1+s) \log \rho - s \log \sigma} \right].
\]
By the Golden-Thompson inequality
\[ \text{Tr} \left[ e^{A+B} \right] \leq \text{Tr} \left[ e^A e^B \right] \]
for Hermitian operators \( A, B \) (see e.g. [1], p. 261), we can see that \( \psi(s) \geq \overline{\psi}(s) \) and that the equality holds if and only if \( \rho \) and \( \sigma \) commute. Thus we have
\[
\min_{\hat{\rho} : D(\hat{\rho}||\sigma) = r} D(\hat{\rho}||\rho) \geq \varphi(\lambda^*). \tag{36}
\]

7. Concluding Remarks

We have derived a new inequality for quantum hypothesis testing problem and proved the quantum Stein’s lemma and the strong converse property as its applications. The method used here is a kind of combination of classical large deviation technique such as Cramér’s theorem and the use of the monotonicity property of some quantum information measures.

It is interesting to compare the results and the argument of this paper with those of [11] which proved the strong converse in the quantum channel coding problem. The former is essentially based on the operator convexity of \( f(u) = u^t \) for \( 1 \leq t \leq 2 \) or \( -1 \leq t \leq 0 \), while the latter relies upon the operator concavity and the operator monotonicity of \( f(u) = u^t \) for \( 0 \leq t \leq 1 \).

In classical hypothesis testing the optimal exponent has been obtained by [7]. On the other hand, it is yet to be determined whether the exponent \( u(r) \) derived here for the quantum hypothesis testing is optimal or not.

Appendix

A. Definition of \( S_f(\rho||\sigma) \)

Here we summarize the original definition of \( S_f(\rho||\sigma) \) given by Petz [14], translating it into our notation, for readers’ convenience. Given arbitrary strictly positive density operators \( \rho, \sigma \in S(\mathcal{H}) \), define the relative modular operator
\[
\Delta = \Delta_{\sigma,\rho} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})
\]
by
\[
\Delta(A) \overset{\text{def}}{=} \sigma A \rho^{-1} \quad \forall A \in \mathcal{L}(\mathcal{H})
\]
or, equivalently, by
\[
\langle \langle \Delta(A), B \rangle \rangle_{\rho} = \langle \langle A, B \rangle \rangle_{\sigma}^+ \quad \forall A, B \in \mathcal{L}(\mathcal{H})
\]
where $\langle \cdot, \cdot \rangle^\rho$ and $\langle \cdot, \cdot \rangle^\sigma$ denote the inner products on $\mathcal{L}(\mathcal{H})$ such that

$$\langle A, B \rangle^\rho \overset{\text{def}}{=} \text{Tr} [\rho A^* B] \quad \text{and} \quad \langle A, B \rangle^\sigma \overset{\text{def}}{=} \text{Tr} [\sigma B A^*].$$

Then since $\Delta$ is Hermitian and positive definite as an operator on the Hilbert space $(\mathcal{L}(\mathcal{H}), \langle \cdot, \cdot \rangle^\rho)$, the ordinary operator calculus enables us to define the Hermitian operator $f(\Delta)$ on the same Hilbert space for an arbitrary function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$. Now let

$$S^f_A (\rho \| \sigma) \overset{\text{def}}{=} \langle A, f(\Delta)(A) \rangle^\rho = \text{Tr} [\rho A^* f(\Delta)(A)]$$

for an arbitrary operator $A \in \mathcal{L}(\mathcal{H})$, and let

$$S_f (\rho \| \sigma) \overset{\text{def}}{=} S^f_I (\rho \| \sigma).$$

Petz called $S^f_A (\rho \| \sigma)$ an quasientropy and proved its monotonicity in the case when $f$ is operator convex. Using the spectral representations (13) we have

$$\Delta(G_j F_i) = \frac{b_j}{a_i} G_j F_i \quad \text{and} \quad f(\Delta)(G_j F_i) = f \left( \frac{b_j}{a_i} \right) G_j F_i$$

which, combined with $I = \sum_i \sum_j G_j F_i$, leads to (14).

**B. Proof of (35)**

Let us define a family of probability distributions $\{p(s) \mid -\infty < s < \infty\}$ including $p(0) = p$ and $p(-1) = q$ as follows:

$$p(s)_j \overset{\text{def}}{=} e^{-\tilde{\psi}(s)} p_j^{1+s} q_j^{-s} \quad (j \in \mathcal{X})$$

where

$$\tilde{\psi}(s) \overset{\text{def}}{=} \log \sum_{j \in \mathcal{X}} p_j^{1+s} q_j^{-s}.$$

On the assumption that $r$ is greater than but sufficiently near $D(p\|q)$, there exists a number $0 < s^* < 1$ such that $D(p(s^*)\|q) = r$. We claim

$$p(s^*) = \arg \min_{\hat{p} : D(\hat{p}\|q) = r} D(\hat{p}\|p). \quad (37)$$
To prove this, let \( \hat{p} \) be an arbitrary distribution satisfying \( D(\hat{p} || q) = r \). Then we have
\[
0 = D(\hat{p} || q) - D(p(s^*) || q) \\
= \sum_j \{ \hat{p}_j \log \hat{p}_j - \hat{p}_j \log q_j - p(s^*)_j \log p(s^*)_j + p(s^*)_j \log q_j \} \\
= D(\hat{p} || p(s^*)) + \sum_j \{ \hat{p}_j - p(s^*)_j \} \{ \log p(s^*)_j - \log q_j \} \\
= D(\hat{p} || p(s^*)) + (1 + s^*) \sum_j \{ \hat{p}_j - p(s^*)_j \} \{ \log p_j - \log q_j \}
\]
and similarly
\[
\Delta \overset{\text{def}}{=} D(\hat{p} || p) - D(p(s^*) || p) = D(\hat{p} || p(s^*)) + s^* \sum_j \{ \hat{p}_j - p(s^*)_j \} \{ \log p_j - \log q_j \}.
\]
Eliminating the common term \( \sum_j \{ \hat{p}_j - p(s^*)_j \} \{ \log p_j - \log q_j \} \) from these equations we obtain
\[
\Delta = \frac{1}{1 + s^*} D(\hat{p} || p(s^*)) \geq 0,
\]
which proves (37). Next we claim
\[
D(p(s^*) || p) = \tilde{\varphi}(\lambda^*) = r - \lambda^*, \tag{38}
\]
where
\[
\tilde{\varphi}(\lambda) \overset{\text{def}}{=} \max_{0 \leq s \leq 1} \left\{ \lambda s - \tilde{\psi}(s) \right\}, \quad \lambda^* \overset{\text{def}}{=} \sum_j p(s^*)_j (\log p_j - \log q_j).
\]
Since a direct calculation based on the assumption that \( D(p(s^*) || q) = r \) yields
\[
D(p(s^*) || p) = \lambda^* s^* - \tilde{\psi}(s^*) = r - \lambda^*,
\]
we have only to show that
\[
s^* = \arg \max_s \left\{ \lambda^* s - \tilde{\psi}(s) \right\}.
\]
Observing that \( \tilde{\psi}(s) \) is a convex function, this equation is reduced to \( \lambda^* = \tilde{\psi}'(s^*) \), which can be verified by a direct calculation (cf. (20)). Now it is easy to see that (35) is obtained from (34), (37), and (38) in the same way as the proof of (29).
Acknowledgments

The authors wish to thank Prof. F. Hiai for his helpful comments. They are also grateful to Prof. T. S. Han for encouraging them to prove the strong converse for quantum hypothesis testing.

References

CHAPTER 3

The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability

Fumio Hiai and Dénes Petz

Abstract. Umegaki’s relative entropy $S(\omega, \varphi) = \text{Tr} D_\omega (\log D_\omega - \log D_\varphi)$ (of states $\omega$ and $\varphi$ with density operators $D_\omega$ and $D_\varphi$, respectively) is shown to be an asymptotic exponent considered from the quantum hypothesis testing viewpoint. It is also proved that some other versions of the relative entropy give rise to the same asymptotics as Umegaki’s one. As a byproduct, the inequality $\text{Tr} A \log AB \geq \text{Tr} A (\log A + \log B)$ is obtained for positive definite matrices $A$ and $B$.

1. Introduction and Motivation

The relative entropy is an information quantity attached to two states of a system. In commutative (or classical) probability theory the states correspond to measures on a measurable space. When $\nu = (\nu_1, \nu_2, \ldots, \nu_n)$ and $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ are probability distributions, for the sake of simplicity, on an n-point space, the relative entropy (called also information divergence) introduced by Kullback and Leibler [17] is defined by

$$S(\nu, \mu) = \sum_i \nu_i \log \frac{\nu_i}{\mu_i}. \quad (1)$$

In noncommutative (or quantum) probability theory the relative entropy of normal positive functionals was first studied by Umegaki [33] in the case of semifinite von Neumann algebras as the noncommutative extension of information divergence. Later on Araki [1, 2] extended it to the case of general von Neumann algebras by means of the notion of a relative modular operator. On the other hand Uhlmann [32] introduced the relative entropy of positive functionals of arbitrary $*$-algebras by a quadratic interpolation method. The importance of relative entropy has been justified by the fact that one encounters this quantity in dealing with a number of different problems.
In quantum theory the states of a system correspond to positive operators of trace one. (These operators are called densities.) In particular, in the setting of matrix algebras Umegaki’s relative entropy of a state $\omega$ with respect to another state $\varphi$ is defined by

$$S(\omega, \varphi) = \text{Tr} D_\omega (\log D_\omega - \log D_\varphi),$$

(2)

where Tr denotes the usual trace on matrices and $D_\omega$ the density of $\omega$ with respect to Tr. [Note that $S(\omega, \varphi)$ sometimes is written as $S(\varphi, \omega)$.] This definition does not seem to be canonical. Compared with (1) one could suppose that other expressions like

$$S_{co}(\omega, \varphi) = \sup \left\{ \sum_i \omega(p_i) \log \frac{\omega(p_i)}{\varphi(p_i)} : p_1, \ldots, p_n \text{ are projectors, } \sum_i p_i = 1 \right\},$$

(3)

$$S_{cp}(\omega, \varphi) = \sup \left\{ \sum_i \omega(a_i) \log \frac{\omega(a_i)}{\varphi(a_i)} : a_1, \ldots, a_n \geq 0, \sum_i a_i = 1 \right\}$$

(4)

or

$$S_{BS}(\omega, \varphi) = \text{Tr} D_\omega \log D_\omega^{1/2} D_\varphi^{-1} D_\omega^{1/2}$$

(5)

are as good as (2). The quantities $S_{co}$ and $S_{cp}$ appeared in [8] and they may be related to observations which are projection-valued or positive operator-valued measures. The definitions (4) and (3) are of the form $\sup \{ S(\omega \circ \alpha, \varphi \circ \alpha) : \alpha \}$, where $\alpha$ runs over all positive (and multiplicative, respectively) unital maps of finite dimensional commutative $C^*$-algebras into the given matrix algebra. Given $a_1, \ldots, a_n \geq 0, \sum_i a_i = 1$, the inequality $\sum_i \omega(a_i) \log \omega(a_i)/\varphi(a_i) \leq S(\omega, \varphi)$ follows by applying the monotonicity of relative entropy [16, 32] to a positive unital map $\alpha((\xi_i)) = \sum_i \xi_i a_i, (\xi_i) \in l^\infty_n$. See also [8, 9] for this inequality. Thus, the inequality $S_{cp}(\omega, \varphi) \leq S(\omega, \varphi)$ holds, while $S_{co}(\omega, \varphi) \leq S(\omega, \varphi)$ is trivial. But the equality here is very restrictive. In fact, it is known [24] that if $\omega$ and $\varphi$ are faithful normal states of a von Neumann algebra $\mathcal{M}$, then $\omega$ must commute with $\varphi$ whenever $S(\omega | \mathcal{N}, \varphi | \mathcal{N}) = S(\omega, \varphi) < +\infty$ holds for some commutative von Neumann subalgebra $\mathcal{N}$ of $\mathcal{M}$. The quantity $S_{BS}$ was introduced in [5] (in a more general setting) and it appeared in [12] in operator form. In this paper it will be shown that the entropy quantities $S_{co}$, $S_{cp}$, and $S$ give rise to the same asymptotic mean in the infinite tensor product system.
We want to deal with the question of the proper definition of information divergence of two states in noncommutative probability theory. This question can be approached from two essentially different points of view. One can search for plausible postulates which should be satisfied by a good notion of relative entropy and one can try to show that $S(\omega, \varphi)$ is the only functional which meets all the desiderata. In this point it was proved in [27] that up to a constant factor only Umegaki’s relative entropy satisfies a reasonable set of postulates. Our approach in the present paper is more pragmatic. We consider the asymptotics of certain probabilities and observe that Umegaki’s relative entropy naturally shows up.

In Sec. 2 of this paper we state the main results in the framework of finite dimensional $C^*$-algebras. Let $\mathcal{A}$ be a finite dimensional $C^*$-algebra (i.e., a finite direct sum of matrix algebras) with a fixed state $\varphi$. As the reference state we take the product state $\varphi_\infty = \otimes_\infty \varphi$ on the infinite $C^*$-tensor product $\mathcal{A}_\infty = \otimes_\infty \mathcal{A}$. Let $\psi$ be a stationary (i.e. invariant for the right shift) state of $\mathcal{A}_\infty$. Then one has the mean relative entropy $S_M(\psi, \varphi_\infty) = \lim_{n \to \infty} \frac{1}{n} S(\psi_n, \varphi_n)$, where $\psi_n = \psi| \otimes_1^n \mathcal{A}$ and $\varphi_n = \varphi| \otimes_1^n \mathcal{A}$.

The mean relative entropy plays an important role in classical as well as quantum statistical mechanics and behaves as a rate function in limit theorems of large deviation type (cf. [11, 25, 28, 29]). Our first theorem says that $\lim_{n \to \infty} \frac{1}{n} S^{\text{co}}(\psi_n, \varphi_n) = S_M(\psi, \varphi)$ for every stationary state $\psi$ of $\mathcal{A}_\infty$. In particular we have $\lim_{n \to \infty} \frac{1}{n} S^{\text{co}}(\omega_n, \varphi_n) = S(\omega, \varphi)$ for every state $\omega$ of $\mathcal{A}$. This has an interesting corollary that $S_{\text{BS}}(\omega, \varphi) \geq S(\omega, \varphi)$ for all states $\omega$ and $\varphi$ of $\mathcal{A}$.

Moreover for $n \geq 1$ and $0 < \varepsilon < 1$ let us introduce the following quantities:

$$\beta_\varepsilon(\psi_n, \varphi_n) = \inf \left\{ \log \varphi_n(q) : q \text{ is a projection in } \otimes_1^n \mathcal{A} \text{ with } \psi_n \geq 1 - \varepsilon \right\}$$

and

$$S_{\text{pr}}(\psi_n, \varphi_n) = \sup \left\{ \psi_n(q) \log \frac{\psi_n(q)}{\varphi_n(q)} + (1 - \psi_n(q)) \log \frac{1 - \psi_n(q)}{1 - \varphi_n(q)} : q \text{ is a projection in } \otimes_1^n \mathcal{A} \right\}.$$ 

The quantity $\beta_\varepsilon(\psi_n, \varphi_n)$ has a natural meaning from the viewpoint of quantum hypothesis testing (cf. [4, 6, 13]). More precisely, let us suppose two hypotheses $H_0$ and $H_1$ so that the system $\mathcal{A}_\infty$ has states $\psi$ and $\varphi_\infty$ under $H_0$ and $H_1$, respectively. A projection $q$ in $\otimes_1^n \mathcal{A}$ means a “quantum question”
of size $n$, whose outcomes are the eigenvalues 1 or 0. We decide that $H_0$ (respectively $H_1$) is true if the outcome of $q$ is 1 (respectively 0). Then $\varphi_n(q)$ (respectively $\varphi_n(1-q)$) gives the “probability” of the error of accepting $H_0$ (respectively $H_1$) when $H_1$, (respectively $H_0$) actually is true. In this way we can consider the quantity $\exp\{\beta_\varepsilon(\psi_n, \varphi_n)\}$ as the bound of the first error probability over all decision rules of size $n$ such that the second error probability does not exceed $\varepsilon$.

Now assume that $\psi$ is completely ergodic in the sense that it is ergodic for any power of the right shift. (This is the case when $\psi$ is weakly mixing.) Then our second theorem says that $\limsup_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) \leq -S_M(\psi, \varphi_\infty)$ and $\liminf_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) \geq -\frac{1}{1-\varepsilon}S_M(\psi, \varphi_\infty)$. Thus we can relate the mean relative entropy to a certain kind of asymptotic error bound in the quantum hypothesis testing. It could be mentioned that a desire for the visualization of noncommutative relative entropy as the logarithm of certain probabilities was formulated in [8] in connection with an interpretation of quantum theory. We finally establish that $\lim_{n \to \infty} \frac{1}{n} S_{pr}(\psi_n, \varphi_n) = S_M(\psi, \varphi_\infty)$.

The proofs of these theorems are given in Sec. 3. The first theorem can be proved by a direct combinatorial computation. The proof of the second theorem is based on the Shannon-McMillan-Breiman theorem and the mean ergodic theorem together with the first theorem.

In Sec. 4 we note that our theorems hold in $AF C^*$-algebras or hyperfinite von Neumann algebras as well. Furthermore we show that if $\psi$ is a tracial ergodic state of $A_\infty$, then $\lim_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) = -S_M(\psi, \varphi_\infty)$ holds for every $0 < \varepsilon < 1$.

2. Main Results in Finite Dimensional $C^*$-Algebras

In this section we state our main theorems in the setting of finite dimensional $C^*$-algebras. Their proofs will be presented in Sec. 3. Although the theorems hold true in the framework of AF $C^*$-algebras or hyperfinite von Neumann algebras as will be noted in Sec. 4, these extensions are very easy and our essential ideas consist in the finite dimensional case; so we restrict our detailed discussions to this case. First let us fix the notations. Let $\mathcal{A}$ be a finite dimensional $C^*$-algebra. Then $\mathcal{A}$ is identified with $\otimes_{l=1}^L M_{d_l}(C)$ which is the direct sum of $d_l \times d_l$ matrix algebras $M_{d_l}(C)$, canonically represented on the Hilbert space $\otimes_{l=1}^L C^{d_l}$. We denote by $\mathcal{A}_n$ the $n$-fold $C^*$-tensor product $\otimes_1^n \mathcal{A}$ for $n \geq 1$, and by $\mathcal{A}_\infty$ the two-sided infinite $C^*$-tensor product $\otimes_{-\infty}^\infty \mathcal{A}$. Then $\{\mathcal{A}_n\}$ is considered as an increasing sequence of (fi-
nite dimensional) $C^*$-subalgebras of $A_\infty$ by the natural inclusions. Let $\gamma$ denote the right shift automorphism of $A_\infty$. A state $\psi$ of $A_\infty$ is said to be stationary if $\psi$ is $\gamma$-invariant (i.e. $\psi \circ \gamma = \varphi$). For a state $\varphi$ of $A_\infty$ we denote by $\varphi_\infty$ the infinite product state $\otimes_{-\infty}^\infty \varphi$, and let $\varphi_n = \varphi_\infty | A_n (= \otimes_1^n \varphi)$. We fix a state $\varphi$ of $A$ in the following discussions.

Let $\psi$ be a stationary state of $A_\infty$ and $\psi_n = \psi| A_n$, $n \geq 1$. The relative entropies $S(\psi_n, \varphi_n)$ are defined as (2). Then in view of the superadditivity of relative entropy [23] and the stationarity of $\psi$ we get

$$S(\psi_{m+n}, \varphi_{m+n}) \geq S(\psi_m, \varphi_m) + S(\psi_n, \varphi_n), \quad m, n \geq 1$$

so that $\lim_{n \to \infty} \frac{1}{n} S(\psi_n, \varphi_n)$ exists; in fact (cf. [10])

$$\lim_{n \to \infty} \frac{1}{n} S(\psi_n, \varphi_n) = \sup_{n \geq 1} \frac{1}{n} S(\psi_n, \varphi_n). \quad (8)$$

We denote this limit by $S_M(\psi, \varphi_\infty)$, which is called the mean relative entropy of $\psi$ with respect to $\varphi_\infty$. Note that if $\omega$ is a state of $A$ then $S_M(\omega_\infty, \varphi_\infty) = S(\omega, \varphi)$ because $S(\omega_n, \varphi_n) = n S(\omega, \varphi)$, $n \geq 1$. In Sec. 1 besides Umegaki’s relative entropy we referred to some other entropy quantities (3)-(5) and (7). The quantities $S_{co}(\psi_n, \varphi_n)$ as in (3) are equivalently defined by

$$S_{co}(\psi_n, \varphi_n) = \sup \{ S(\psi_n|B, \varphi_n|B) : B \text{ is a commutative } C^*-\text{subalgebra of } A_n \}.$$ 

The next theorem shows that the asymptotic limit of $S_{co}(\psi_n, \varphi_n)$ exists and coincides with the mean relative entropy.

**Theorem 1:** For every stationary state $\psi$ of $A_\infty$,

$$\lim_{n \to \infty} \frac{1}{n} S_{co}(\psi_n, \varphi_n) = S_M(\psi, \varphi_n). \quad (9)$$

A stationary state $\psi$ of $A$ is said to be ergodic if it is extremal in the set of stationary states. For ergodicity in general $C^*$-dynamical systems, see [10, 30] for instance. We say that $\psi$ is weakly mixing if for every $a, b \in A_\infty$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\psi(\gamma^i(a)b) - \psi(a)\psi(b)| = 0.$$

Obviously this is the case when $\psi$ is strongly mixing (or strongly clustering), that is, $\lim_{n \to \infty} \psi(\gamma^n(a)b) = \psi(a)\psi(b)$ for every $a, b \in A_\infty$. As in classical ergodic theory it is known [10] that if $\psi$ is weakly mixing then it is ergodic. Note that the product state $\omega_\infty$ defined by a state $\omega$ of $A$ is strongly mixing.
In the following we say that $\psi$ is completely ergodic if it is ergodic for all $\gamma^n$, $n \geq 1$. We know that a weakly mixing state $\psi$ is completely ergodic because it is weakly mixing for all $\gamma^n$.

Let $\beta_n(\psi_n, \varphi_n)$ be defined by (6) for $n \geq 1$ and $0 < \varepsilon < 1$. The next theorem shows that we have for large $n$

$$\exp \left\{ \frac{1}{n} \beta_n(\psi_n, \varphi_n) \right\} \approx \exp \{-S_M(\psi, \varphi_\infty)\}$$

when $\varepsilon$ is sufficiently small and $\psi$ is completely ergodic. Thus $\exp\{-S_M(\psi, \varphi_\infty)\}$ can be considered as the asymptotic error bound in the quantum hypothesis test for $\{\psi, \varphi_\infty\}$.

**Theorem 2:** If $\psi$ is a completely ergodic state of $A_\infty$, then for every $0 < \varepsilon < 1$,

$$\limsup_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) \leq -S_M(\psi, \varphi_\infty),$$

$$\liminf_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) \geq -\frac{1}{1-\varepsilon} S_M(\psi, \varphi_\infty).$$

It may be possible that $\lim_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) = -S_M(\psi, \varphi_\infty)$ for every $0 < \varepsilon < 1$ and every ergodic state $\psi$. In particular when $\psi$ is a tracial ergodic state, this will be shown in Sec. 4.

The quantities $S_{pr}(\psi_n, \varphi_n)$ in (7) are also defined by

$$S_{pr}(\psi_n, \varphi_n) = \sup \{S(\psi_n|B, \varphi_n|B) : \}$$

$B$ is a two dimensional subalgebra of $A_n$.

Note that

$$S_{pr}(\psi_n, \varphi_n) \leq S_{co}(\psi_n, \varphi_n) \leq S_{cp}(\psi_n, \varphi_n) \leq S(\psi_n, \varphi_n)$$

by the monotonicity of relative entropy as mentioned in Sec. 1.

As for completely ergodic states we can make Theorem 1 extremely sharp as follows. Indeed the method in proving (11) will work for Theorem 3 as well.

**Theorem 3:** If $\psi$ is a completely ergodic state of $A_\infty$, then

$$\lim_{n \to \infty} \frac{1}{n} S_{pr}(\psi_n, \varphi_n) = S_{M}(\psi, \varphi_\infty).$$

As a special case we have for every state $\omega$ of $A$,

$$\lim_{n \to \infty} \frac{1}{n} S_{pr}(\omega_n, \varphi_n) = \lim_{n \to \infty} \frac{1}{n} S_{co}(\omega_n, \varphi_n) = \lim_{n \to \infty} \frac{1}{n} S_{cp}(\omega_n, \varphi_n) = S(\omega, \varphi).$$

(13)
This means that Umegaki’s relative entropy comes out when we first adopt any of the quantities $S_{\text{pr}}$, $S_{\text{co}}$ or $S_{\text{cp}}$ and then take asymptotics.

The following examples show that the ergodicity assumption of $\psi$ is essential in Theorems 2 and 3. We are indebted to the referee for the first example.

Example 4: (1) Let $0 < \varepsilon < 1/2$ and $\psi$ be a stationary state of $A_{\infty}$ with $\psi \neq \varphi_{\infty}$. Then $S_M(2\varepsilon \varphi_{\infty} + (1 - 2\varepsilon)\psi, \varphi_{\infty}) > 0$ because otherwise by (8) we get $2\varepsilon \varphi_{\infty} + (1 - 2\varepsilon)\psi = \varphi_{\infty}$ so that $\psi = \varphi_{\infty}$. On the other hand, since $2\varepsilon \varphi_n(q) + (1 - 2\varepsilon)\psi_n(q) \geq 1 - \varepsilon$ implies $\varphi_n(q) \geq 1/2$ for a projection $q$ in $A_n$, we have

$$\beta_\varepsilon(2\varepsilon \varphi_n + (1 - 2\varepsilon)\psi_n, \varphi_n) \geq -\log 2.$$ 

Hence

$$\lim_{n \to \infty} \frac{1}{n} \beta_\varepsilon(2\varepsilon \varphi_n + (1 - 2\varepsilon)\psi_n, \varphi_n) = 0.$$ 

(2) Let $A = C \oplus C$ and $\omega^1$, $\omega^2$, $\varphi$ be given with the densities $(1, 0)$, $(0, 1)$, $(\alpha, 1 - \alpha)$, respectively, where $0 < \alpha < 1/2$. Let $\psi = \frac{1}{2}(\omega_{\infty}^1 + \omega_{\infty}^2)$. By the affinity of $S_M(\cdot, \varphi_{\infty})$ (see [28]) we get

$$S_M(\psi, \varphi_{\infty}) = \frac{1}{2}\{S(\omega^1, \varphi) + S(\omega^2, \varphi)\} = -\frac{1}{2}\{\log \alpha + \log(1 - \alpha)\}.$$ 

But we easily see that $S_{\text{pr}}(\psi_n, \varphi_n)$ is the maximum of

$$\log \frac{1}{\alpha^n + (1 - \alpha)^n} = -n \log \{\alpha^n + (1 - \alpha)^n\}^{1/n}$$ 

and

$$\frac{1}{2} \log \frac{1}{2\alpha^n} + \frac{1}{2} \log \frac{1}{2(1 - \alpha)^n} = \max \left\{-\log(1 - \alpha), -\frac{1}{2} \log \alpha\right\} < S_M(\psi, \varphi_{\infty}).$$

In the rest of this section, using Theorem 1 let us establish the relation between the relative entropy $S(\omega, \varphi)$ and the entropy quantity $S_{\text{BS}}(\omega, \varphi)$ in (5). The expression (5) implicitly means that $S_{\text{BS}}(\omega, \varphi) = +\infty$ if the support projection of $\omega$ is not dominated by that of $\varphi$. The main properties of $S_{\text{BS}}$ follow from [12] devoted to an operator-valued relative entropy. We here state the additivity and the monotonicity of $S_{\text{BS}}$.

Proposition 5: (1) $S_{\text{BS}}(\omega_1 \otimes \omega_2, \varphi_1 \otimes \varphi_2) = S_{\text{BS}}(\omega_1, \varphi_1)$ when $\omega_i$ and $\varphi_i$ are states of (finite dimensional) $C^*$-algebras $A_i$, $i = 1, 2$.

(2) $S_{\text{BS}}(\omega|B, \varphi|B) \leq S_{\text{BS}}(\omega, \varphi)$ for any $C^*$-subalgebra $B$ of $A$. 
Indeed (1) is obvious from the definition. Although (2) follows from the operator-valued version of [12], we briefly recall the proof for the convenience of the reader. We may assume that $\varphi$ is faithful. Since $X(\log XX^*)X = (\log XX^*)X$ holds for a matrix $X$, it follows that
\[
S_{BS}(\omega, \varphi) = -\text{Tr}(D_\varphi \eta(D_\varphi^{-1/2}D_\omega D_\varphi^{-1/2}))
\]
where $\eta(t) = -t \log t$, $t \geq 0$. By the operator-concavity of $\eta$ we get
\[
\alpha(\eta(D_\varphi^{-1/2}D_\omega D_\varphi^{-1/2})) \leq \eta(\alpha(D_\varphi^{-1/2}D_\omega D_\varphi^{-1/2}))
\]
where $\alpha(X) = E(D_\varphi)^{-1/2}E(D_\varphi^{1/2}XD_\varphi^{1/2})E(D_\varphi)^{-1/2}$ with the conditional expectation $E$ from $\mathcal{A}$ onto $\mathcal{B}$ with respect to Tr. Hence
\[
E(D_\varphi^{1/2}(D_\varphi^{-1/2}D_\omega D_\varphi^{-1/2})D_\varphi^{1/2}) \\ \leq E(D_\varphi)^{1/2}E(D_\omega)E(D_\varphi)^{-1/2}E(D_\varphi)^{1/2}.
\]
Taking the trace of both sides proves (2).

Now Theorem 1 together with Proposition 5 has an interesting consequence as follows.

**Corollary 6:** $S_{BS}(\omega, \varphi) \geq S(\omega, \varphi)$ for all states $\omega$ and $\varphi$ of $\mathcal{A}$.

**Proof:** Since $S_{BS} = S$ for commuting states, we have $S_{BS}(\omega, \varphi) \geq S_{co}(\omega, \varphi)$ by (2) of the above proposition. In particular, for every $n \geq 1$,
\[
S_{BS}(\omega_n, \varphi_n) \geq S_{co}(\omega_n, \varphi_n),
\]
so that by the above (1),
\[
S_{BS}(\omega, \varphi) \geq \frac{1}{n}S_{co}(\omega_n, \varphi_n).
\]
Letting $n \to \infty$ we infer the corollary due to Theorem 1. \[\blacksquare\]

It is quite remarkable that the corollary is equivalent to the trace inequality
\[
\text{Tr } A \log A^{1/2}BA^{1/2} \geq \text{Tr } A(\log A + \log B) \tag{14}
\]
for nonnegative matrices $A$ and $B$. In fact, (14) is immediate from the corollary when $B$ is invertible. Take the limit from $B + \varepsilon I$, $\varepsilon > 0$, for general nonnegative $B$. When $A$ and $B$ are positive invertible, one can define $\log AB$ by analytical functional calculus or by power series and get the equality
\[
\text{Tr } A \log A^{1/2}BA^{1/2} = \text{Tr } A \log AB
\]
because $\text{Tr } A(A^{1/2}BA^{1/2})^n = \text{Tr } A(AB)^n$ for $n \geq 1$. 
3. Proofs of Theorems

In this section we present the proofs of Theorems 1–3. Let us keep the notations fixed in the previous section. Let \( \text{Tr} \) denote the canonical trace of \( A \) such that \( \text{Tr}(e) = 1 \) for all one dimensional projections \( e \) in \( A \). Let \( D_\varphi \), be the density of \( \varphi \) with respect to \( \text{Tr} \), and \( K \) be the sum of the sizes of simple summands of \( A \) (i.e. \( K = \sum_{i=1}^L d_i \)). Taking the spectral decomposition of \( D_\varphi \), one can write \( D_\varphi = \sum_{k=1}^K \lambda_k e_k \), where \( e_k \) are one dimensional projections. Let \( n \) be an arbitrary fixed positive integer. For each \( K \)-tuple \((n_1, n_2, \ldots, n_K)\) of nonnegative integers with \( \sum_{k=1}^K n_k = n \), we denote by \( I_{n_1,\ldots,n_K} \) the set of all \((i_1, i_2, \ldots, i_n)\) such that \#\([j : i_j = k]\) = \( n_k \) for \( 1 \leq k \leq K \), and define the projection \( p_{n_1,\ldots,n_K} \) in \( A \) by

\[
p_{n_1,\ldots,n_K} = \sum_{(i_1,\ldots,i_n)\in I_{n_1,\ldots,n_K}} e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n}.
\]

Then \( \sum_{n_1,\ldots,n_K} p_{n_1,\ldots,n_K} = 1 \) and the density \( D_{\varphi_n} \) of \( \varphi_n \) with respect to \( \text{Tr}_n = \bigotimes_i^n \text{Tr} \) is given by

\[
D_{\varphi_n} = \bigotimes_i^n D_\varphi = \sum_{n_1,\ldots,n_K} \left( \prod_{k=1}^K \lambda_k^{n_k} \right) p_{n_1,\ldots,n_K}. \tag{15}
\]

Let \( E_n \) denote the conditional expectation from \( A_n \) onto \( \bigotimes_{n_1,\ldots,n_K} p_{n_1,\ldots,n_K} \) \( A_n p_{n_1,\ldots,n_K} \) with respect to \( \text{Tr}_n \), which is given by

\[
E_n(x) = \sum_{n_1,\ldots,n_K} p_{n_1,\ldots,n_K}xp_{n_1,\ldots,n_K}, \quad x \in A_n. \tag{16}
\]

The (von Neumann) entropy \( S(\omega) \) of a state \( \omega \) of \( A_n \) is defined by \( S(\omega) = -\text{Tr}_n(D_\omega \log D_\omega), \) where \( D_\omega \) is the density of \( \omega \) with respect to \( \text{Tr}_n \).

Lemma 7: If \( \omega \) is a state of \( A_n \) and \( B \) is the commutative subalgebra of \( A_n \) generated by \( \{p_{n_1,\ldots,n_K}D_\omega p_{n_1,\ldots,n_K}\}_{n_1,\ldots,n_K} \cup \{p_{n_1,\ldots,n_K}\}_{n_1,\ldots,n_K} \), then

\[
S(\omega, \varphi_n) = S(\omega|B, \varphi|B) + S(\omega \circ E_n) - S(\omega).
\]

Proof: Let \( s(\omega) \) and \( s(\varphi_n) \) denote the support projections of \( \omega \) and \( \varphi_n \), respectively. Since \( D_{\varphi_n} \in B \) by (15), we have \( s(\varphi_n) = s(\varphi_n|B) \in B \). If \( s(\omega) \leq s(\varphi_n) \) is not satisfied, then the desired equality holds because \( S(\omega, \varphi_n) = S(\omega|B, \varphi_n|B) = +\infty \). So suppose \( s(\omega) \leq s(\varphi_n) \). In this case we may assume that \( \varphi_n \) is faithful; otherwise consider the restrictions of \( \omega \) and \( \varphi_n \) to \( s(\varphi_n)A_n s(\varphi_n) \) and \( Bs(\varphi_n) \). Since \( B \) is included in the centralizer...
of \( \varphi_n \), the conditional expectation \( E_B \) from \( \mathcal{A}_n \) onto \( \mathcal{B} \) with respect to \( \varphi_n \) exists due to [31]. Hence we get by [14, 22, 23]
\[
S(\omega, \varphi_n) = S(\omega|\mathcal{B}, \varphi|\mathcal{B}) + S(\omega, \omega \circ E_B).
\] (17)
Since \( E_n(D_\omega) \in \mathcal{B} \) and \( \mathcal{B} \subset E_n(\mathcal{A}_n) \) by (16), we get for every \( a \in \mathcal{A}_n \),
\[
(\omega \circ E_B)(a) = \text{Tr}_n(D_\omega E_B(a)) = \text{Tr}_n(E_n(D_\omega)E_B(a))
\]
\[
= \varphi_n(D_{\varphi_n}^{-1}E_n(D_\omega)E_B(a)) = \varphi_n(E_B(D_{\varphi_n}^{-1}E_n(D_\omega)a))
\]
\[
= \varphi_n(D_{\varphi_n}^{-1}E_n(D_\omega)a) = \text{Tr}_n(E_n(D_\omega)a),
\]
so that \( E_n(D_\omega) \) is the density of \( \omega \circ E_B \) as well as \( \omega \circ E_n \) with respect to \( \text{Tr}_n \). Therefore \( \omega \circ E_B = \omega \circ E_n \) and
\[
S(\omega, \omega \circ E_n) = \text{Tr}_n D_\omega (\log D_\omega - \log E_n(D_\omega)) = S(\omega \circ E_n) - S(\omega).
\] (18)
The desired equality follows from (17) and (18).

**Lemma 8:** For every state \( \omega \) of \( \mathcal{A}_n \),
\[
S(\omega \circ E_n) - S(\omega) \leq K \log(n + 1).
\]

**Proof:** In view of the joint convexity of the relative entropy [1, 16] it suffices by (18) to show the case when \( \omega \) is a pure state. In this case \( D_\omega \) is a one dimensional projection and \( S(\omega) = 0 \). Since each \( p_{n_1,\ldots,n_K}D_\omega p_{n_1,\ldots,n_K} \) is of rank one or zero, it follows that the rank of \( E_n(D_\omega) \) is at most
\[
\# \left\{ (n_1,\ldots,n_k) : n_1,\ldots,n_k \geq 0, \sum_{k=1}^K n_k = k \right\} \leq (n + 1)^K.
\]
Therefore
\[
S(\omega \circ E_n) \leq \log(n + 1)^K = K \log(n + 1),
\]
as desired.

**Proof of Theorem 1:** For every \( n \) we have
\[
S_{\text{co}}(\psi_n, \varphi_n) \leq S(\psi_n, \varphi_n) \leq S_{\text{co}}(\psi_n, \varphi_n) + K \log(n + 1),
\]
by (12) and by Lemmas 7 and 8 applied to \( \psi_n \). This proves (9).

In the sequel of this section we assume that \( \psi \) is a completely ergodic state of \( \mathcal{A}_\infty \).

**Proof of (10) of Theorem 2:** For any \( r > -S_M(\psi, \varphi_\infty) \) let us choose \( h < S_M(\psi, \varphi_\infty) \) and \( \delta > 0 \) with \( -h + \delta < r \). By Theorem 1 there exists a commutative \( C^* \)-subalgebra of \( \mathcal{B} \) of \( \mathcal{A}_l \) for some \( l \geq 1 \) such that
\[
S(\psi_l|\mathcal{B}, \varphi_l|\mathcal{B}) \geq lh.
\] (19)
We can consider $B_k$ as a $C^*$-subalgebra of $A_{kl}$, $k \geq 1$, and $B_\infty$ as a $C^*$-subalgebra of $A_\infty$. Let $\sigma = \gamma^i|B_\infty$, the right shift on $B_\infty$. Define $\mu = \varphi_i|B$, $\nu = \psi|B_\infty$, $\mu_\infty = \otimes^\infty_\infty \mu$, $\mu_k = \mu_\infty|B_k$ and $\nu_k = \nu|B_k$. These states may be identified with the probability measures on the corresponding underlying spaces. Since $\psi$ is completely ergodic, we can readily see that $\nu$ is ergodic for $\sigma$. In the following we work in the (commutative) von Neumann algebra $\pi_\nu(B_\infty)^\prime\prime$ where $\pi_\nu$ is the GNS representation of $B_\infty$ associated with $\nu$. We denote the normal extensions of $\nu$ and $\sigma$ to $\pi_\nu(B_\infty)^\prime\prime$ by the same $\nu$ and $\sigma$.

First suppose $\nu_1 \ll \mu$ (i.e. $s(\nu_1) \leq s(\mu)$) is not satisfied. Then there is a projection $p$ in $B$ such that $\mu(p) = 0$ and $\nu_1(p) > 0$. Since the ergodicity of $\nu$ implies that $\pi_\nu \left( \bigvee_{i=0}^{k-1} \sigma^i(p) \right) \to 1$ strongly as $k \to \infty$, there exists $k_0$ such that $\nu \left( \bigvee_{i=0}^{k_0-1} \sigma^i(p) \right) \geq 1 - \varepsilon$. Set $q = \bigvee_{i=0}^{k_0-1} \sigma^i(p)$ and $n_0 = k_0 l$. Then $q$ is a projection in $A_{n_0}$ such that $\varphi(q) \geq 1 - \varepsilon$ and

$$
\varphi_{n_0}(q) = \mu_{k_0} \left( \bigvee_{i=0}^{k_0-1} \sigma^i(p) \right) \leq k_0 \mu(p) = 0.
$$

Therefore $\beta_\varepsilon(\psi_n, \varphi_n) = -\infty$ for all $n \geq n_0$, proving (10).

Next suppose $\nu_1 \ll \mu$. Then we see that $\nu_k \ll \mu_k$ for every $k$ because $s(\nu_k) \leq \otimes^k_1 s(\nu_1) \leq \otimes^k_1 s(\mu) = s(\mu_k)$. Let $m$ denote the trace of $B$ such that $m(e) = 1$ for all atoms $e$ in $B$ (i.e. $m$ is the counting measure on the underlying space of $B$). Let us consider the selfadjoint operators $H_k$, $k \geq 1$, in $\pi_\nu(B_\infty)^\prime\prime$ given by

$$
H_k = \frac{1}{k} \left\{ \pi_\nu \left( \log \frac{d\nu_k}{dm_k} \right) - \pi_\nu \left( \log \frac{d\mu_k}{dm_k} \right) \right\}
= \frac{1}{k} \pi_\nu \left( \log \frac{d\nu_k}{dm_k} \right) - \frac{1}{k} \sum_{i=0}^{k-1} \sigma^i \left( \pi_\nu \left( \log \frac{d\mu}{dm} \right) \right). \tag{20}
$$

By the Shannon-McMillan-Breiman theorem (cf. [3, 21]) the first term of (20) converges $\nu$-almost uniformly as $k \to \infty$ to the Kolmogorov-Sinai entropy $h_\nu(\sigma)$ of $\sigma$ relative to $\nu$. On the other hand, by the mean ergodic theorem the second term of (20) converges strongly as $k \to \infty$ to $\nu_1(\log d\mu/dm)$. Thus we see that $H_k$ converges in $\mu$-measure to

$$
h_0 = \lim_{k \to \infty} \frac{1}{k} \nu_k \left( \log \frac{d\nu_k}{dm_k} - \log \frac{d\mu_k}{dm_k} \right) = \lim_{k \to \infty} \frac{1}{k} S(\nu_k, \mu_k) = S_M(\nu, \mu_\infty).
$$

Now for each $k$ let $p_k$ be the projection in $B_k$ with $p_k \leq s(\nu_k)$ such that $\pi_\nu(p_k)$ is the spectral projection of $H_k$ corresponding to the interval $(h_0 - \delta, h_0 + \delta)$. Then there exists $k_0$ such that $\nu(p_k) \geq 1 - \varepsilon$ for all
$k \geq k_0$. Since $\pi_\nu(p_k)$ is the spectral projection of $\exp(kH_k)$ corresponding to $(e^{k(h_0 - \delta)}, e^{k(h_0 + \delta)})$, we get
\[
\pi_\nu(p_k) \leq e^{-k(h_0 - \delta)} \pi_\nu(p_k) \exp(kH_k) = e^{-k(h_0 - \delta)} \left( p_k \left( \frac{d\nu_k}{dm_k} \right) \left( \frac{d\mu_k}{dm_k} \right)^{-1} \right),
\]
so that since $p_k \leq s(\nu_k)$
\[
p_k \leq e^{-k(h_0 - \delta)} p_k \left( \frac{d\nu_k}{dm_k} \right) \left( \frac{d\mu_k}{dm_k} \right)^{-1}. \tag{22}
\]
Therefore
\[
\mu_k(p_k) \leq e^{-k(h_0 - \delta)} \mu_k \left( p_k \left( \frac{d\nu_k}{dm_k} \right) \right) = e^{-k(h_0 - \delta)} \nu_k(p_k)
\leq e^{-k(h_0 - \delta)} \leq e^{-k(h - \delta)}, \tag{23}
\]
because $h_0 \geq S(\nu_1, \mu) \geq lh$ by (21), (8) and (19). For each $n \geq k_0l$ let $n = kl + j$, where $k \geq k_0$, $0 \leq j < l$, and put $q_n = p_k$. Then we have $q_n \in \mathcal{A}_n$ and $\psi(q_n) = 1 - \varepsilon$, so that
\[
\frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) \leq \frac{1}{n} \log \varphi_n(q_n) \leq \frac{1}{(k+1)l} \log \mu_k(p_k) \leq -\frac{k}{k+1} h + \delta
\]
by (23). This proves (10) thanks to $-h + \delta < r$. \hfill \blacksquare

Proof of (11) of Theorem 2: First suppose $S_M(\psi, \varphi_\infty) = 0$. Then $\psi = \varphi_\infty$ by (2.1) and hence $\beta_\varepsilon(\psi_n, \varphi_n) \geq \log(1 - \varepsilon)$, $n \geq 1$, so that (11) is immediate. Now suppose $S_M(\psi, \varphi_\infty) > 0$. For each $n$ set $\beta_n = \beta_\varepsilon(\psi_n, \varphi_n)$ and choose a projection $q_n$ in $\mathcal{A}_n$ such that $\psi_n(q_n) \geq 1 - \varepsilon$ and $\varphi_n(q_n) < e^{\beta_n + 1}$. Letting $0 < h < S_M(\psi, \varphi_\infty)$, by (10) we have $\beta_n < -nh$ for sufficiently large $n$. Hence $e^{\beta_n + 1} \to 0$ as $n \to \infty$. Define $\mathcal{B}_n = Cq_n + C(1 - q_n)$ which is a two dimensional subalgebra of $\mathcal{A}_n$. Then by monotonicity
\[
S(\psi_n, \varphi_n) = S(\psi_n|\mathcal{B}_n, \varphi_n|\mathcal{B}_n) = F(\psi_n(q_n), \varphi_n(q_n)),
\]
where
\[
F(s, t) = s \log \frac{s}{t} + (1 - s) \log \frac{1 - s}{1 - t}. \tag{24}
\]
Since for $0 < t < s \leq 1$,
\[
\frac{\partial F(s, t)}{\partial t} = \frac{t - s}{t(1 - t)} < 0,
\]
we have for every \( n \) large enough
\[
S(\psi_n, \varphi_n) \geq F(\psi_n(q_n), e^{\beta_n + 1}) \\
= -\psi_n(q_n)(\beta_n + 1) - (1 - \psi_n(q_n)) \log(1 - e^{\beta_n + 1}) \\
+ \psi_n(q_n) \log \psi_n(q_n) + (1 - \psi_n(q_n)) \log(1 - \psi_n(q_n)) \\
\geq -(1 - \varepsilon)(\beta_n + 1) - \log 2.
\]
Therefore
\[
(1 - \varepsilon) \liminf_{n \to \infty} \frac{1}{n} \beta_n \geq -(1 - \varepsilon) \liminf_{n \to \infty} \frac{1}{n} \beta_n \geq -S_M(\psi, \varphi_\infty),
\]
as desired.

**Proof of Theorem 3:** It suffices by monotonicity to show that
\[
\liminf_{n \to \infty} \frac{1}{n} S_{pr}(\psi_n, \varphi_n) \geq S_M(\psi, \varphi_\infty).
\]
We may suppose \( S_M(\omega, \varphi_\infty) > 0 \). Let \( 0 < \varepsilon < 1 \) and \( 0 < h < S_M(\psi, \varphi_\infty) \). By (10) we have \( \beta_\varepsilon(\psi_n, \varphi_n) < -nh \) for sufficiently large \( n \). Hence for each such \( n \) we can choose a projection \( q_n \) in \( A_n \) such that \( \psi_n(q_n) \geq 1 - \varepsilon \) and \( \varphi_n(q_n) < e^{-nh} \). Let \( B_n = Cq_n + C(1 - q_n) \) and \( F \) be the function in (24). Then we have as in the proof of (11),
\[
S_{pr}(\psi_n, \varphi_n) \geq S(\psi_n|B_n, \varphi_n|B_n) \geq F(\psi_n(q_n), \varphi_m(q_n)) \\
\geq F(\psi_n(q_n), e^{-nh}) \geq n(1 - \varepsilon)h - \log 2
\]
for every \( n \) large enough. Therefore
\[
\liminf_{n \to \infty} \frac{1}{n} S_{pr}(\psi_n, \varphi_n) \geq (1 - \varepsilon)h.
\]
We obtain (25) letting \( \varepsilon \to 0 \) and \( h \to S_M(\psi, \varphi_\infty) \).

Indeed it is enough in Theorems 2 and 3 to assume that \( \psi \) is a stationary state of \( A_\infty \) which is ergodic for \( \gamma^n \) for infinitely many \( n \).

## 4. Extensions

In this section we observe that our theorems in Sec. 2 remain true in AF \( C^* \)-algebras or hyperfinite von Neumann algebras. When \( A \) is a general \( C^* \)-algebra (always assumed to be unital), given two states \( \omega \) and \( \varphi \) of \( A \) one can define the relative entropy \( S(\omega, \varphi) \) of \( \omega \) with respect to \( \varphi \) as Uhlmann’s relative entropy [32]. But this is also defined through Araki’s one [1, 2] for normal states of von Neumann algebras as follows: if \( \pi \) is a representation of \( A \) such that \( \omega \) and \( \varphi \) have the respective normal extensions \( \bar{\omega} \) and \( \bar{\varphi} \) to
Fumio Hiai and Dénes Petz

\[ \pi(A)^{\prime\prime} \text{ with } \omega \circ \pi = \omega \text{ and } \varphi \circ \varphi = \varphi, \text{ then we have } S(\omega, \varphi) = S(\omega, \varphi) \text{ (see } [15, 25]). \]

Let \( A \) be a C*-algebra with a fixed state \( \varphi \). As in Sec. 2 we take the C*-tensor products \( A_\infty = \otimes_\infty A, A_n = \otimes_1^n A, n \geq 1 \), the right shift automorphism of \( A_\infty \) and the product state \( \varphi_\infty \) of \( A_\infty \). Moreover for a stationary state \( \psi \) of \( A_\infty \) the mean relative entropy \( S_M(\psi, \varphi_\infty) \) is defined as (8). To extend the theorems to the case of an AF C*-algebra, we first give the following lemma.

Lemma 9: Let \( \{A(j)\} \) be an increasing net of C*-subalgebras of \( A \) such that \( A = \bigcup_j A(j) \). Let \( \psi \) be a stationary state of \( A_\infty \). If \( \varphi(j) = \varphi|A(j) \) and \( \psi(j) = \psi|A(j) \infty \), then \( S_M(\psi(j), \varphi(j)_\infty) \) increases to \( S_M(\psi, \varphi_\infty) \).

Proof: For each \( n \), since \( A_n = \bigcup_j A(j) \), we have

\[
\sup_j S(\psi_n|A(j)_n, \varphi_n|A(j)_n) = S(\psi_n, \varphi_n)
\]

by the martingale convergence of relative entropy [2, 16] applied under some representation of \( A_n \). Hence

\[
S_M(\psi, \varphi_\infty) = \sup_{n \geq 1} \frac{1}{n} S(\psi_n, \varphi_n) = \sup_j \sup_{n \geq 1} \frac{1}{n} S(\psi_n|A(j)_n, \varphi_n|A(j)_n)
\]

\[
= \sup_j S_M(\psi(j), \varphi(j)_\infty).
\]

The increasingness is immediate from the monotonicity of relative entropy.

Now assume that \( A \) is an AF C*-algebra. Then there is an increasing net \( \{A(j)\} \) of finite dimensional subalgebras of \( A \) such that \( A = \bigcup_j A(j) \). Let \( \psi \) be a stationary state of \( A_\infty \) and define \( S_M(\psi_n, \varphi_n), n \geq 1 \), as (3). For each \( j \) and \( n \) we have by Lemmas 7 and 8 applied to \( \psi_n|A(j)_n \),

\[
S(\psi_n|A(j)_n, \varphi_n|A(j)_n) \leq S_M(\psi_n, \varphi_n) + K_j \log(n + 1),
\]

where \( K_j \) is the sum of the sizes of simple summands of \( A(j) \). Hence for every \( j \),

\[
S_M(\psi(j), \varphi(j)_\infty) \leq \lim_{n \to \infty} \frac{1}{n} S_M(\psi_n, \varphi_n),
\]

so that Lemma 9 shows the equality (9). Furthermore by the above argument we can choose, given \( h < S_M(\psi, \varphi_\infty) \), a finite dimensional commutative subalgebra \( B \) of \( A_l \) for some \( l > 1 \) such that (19) holds. Hence the proof of (10) of Theorem 2 works well. Thus we infer that Theorems 1–3
remain true for every stationary or completely ergodic state $\psi$ of $A_\infty$ when $A$ is an AF $C^*$-algebra.

Our theorems can be formulated in the framework of von Neumann algebras too. Let $M$ be a von Neumann algebra with a fixed normal state $\varphi$. Let $M_n$ be the $n$-fold von Neumann tensor product $\otimes^n M$ for $n \geq 1$, and $M_\infty$ the $C^*$-completion of $\bigcup_{n=1}^\infty (\otimes^n M)$. Then $\{M_n\}$ is an increasing sequence of von Neumann algebras included in $M_\infty$. We have the right shift and the product state $\varphi_\infty$ of $M_\infty$ as before. If $\{M(j)\}$ is an increasing net of von Neumann subalgebras of $M$ such that $M = \left(\bigcup_j M(j)\right)''$ and if $\varphi$ is a stationary state of $M_\infty$ such that $\varphi_n = \varphi|M_n$ is normal for every $n$, then the same conclusion as Lemma 9 holds. Now assume that $M$ is hyperfinite (i.e. approximately finite dimensional). Of course this is the case when $M = B(H)$, the algebra of all bounded operators on a Hilbert space $H$. Then the same results as Theorems 1–3 hold for every stationary or completely ergodic state $\varphi$ of $M_\infty$ such that $\varphi_n$ is normal for every $n$. In particular we have (13) for every normal state $\omega$ of $M$.

Assume again that $A$ is finite dimensional with a state $\varphi$. We finally consider the special setting where $\psi$ is a tracial ergodic state of $A_\infty$. In this case we denote by $\tilde{\psi}$ and $\tilde{\gamma}$ the respective normal extensions of $\psi$ and $\gamma$ to $\pi_{\psi}(A_\infty)''$. Then $\tilde{\psi}$ becomes a faithful normal tracial state of $\pi_{\psi}(A_\infty)''$ with $\tilde{\psi} \circ \tilde{\gamma} = \tilde{\psi}$. Let $\text{Tr}$ be the canonical trace of $A$ mentioned at the beginning of Sec. 3 and $Z(A)$ be the center of $A$. Although it is a challenging open problem to establish the noncommutative Shannon-McMillan-Breiman theorem for general ergodic states of $A_\infty$, the special case of the following lemma was given in [20]. In fact, this follows from the classical case for $\psi| \otimes_1^n Z(A)$ because $d\psi_n/d\text{Tr}_n$ belongs to $\otimes_1^n Z(A)$, the center of $A_n$.

**Lemma 10:** Let $A$ and $\psi$ be as above. Then $-\frac{1}{n} \pi_{\psi}(\log d\psi_n/d\text{Tr}_n)$ converges $\tilde{\psi}$-almost uniformly and in $L^1(\tilde{\psi})$-norm to a constant $h$. ($h$ is the dynamical entropy [7] of $\gamma$ relative to $\psi$.)

On the other hand the noncommutative mean ergodic theorem was given in [18]. Based on these convergence results we have the next proposition which is a stronger form of Theorem 2 and extends the classical results in [4,6]. A similar result is given in [19] when $\psi$ is the trace $\text{Tr}$ and $\psi$ is a product state.

**Proposition 11:** Let $\psi$ be a tracial ergodic state of $A_\infty$. Then for every
0 < \varepsilon < 1,
\lim_{n \to \infty} \frac{1}{n} \beta_n(\psi_n, \varphi_n) = -S_M(\psi, \varphi_\infty).

Proof: We shall show that
\lim inf_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) \geq -S_M(\psi, \varphi_\infty). \tag{26}

Our method in the following will allow the proof of (10) of Theorem 2 to be adapted to give
\lim sup_{n \to \infty} \frac{1}{n} \beta_\varepsilon(\psi_n, \varphi_n) \leq -S_M(\psi, \varphi_\infty).

When \( s(\psi_1) \leq s(\varphi) \) is not satisfied, we can see arguing as in the proof of (10) that \( \beta_\varepsilon(\psi_n, \varphi_n) = -\infty \) for every \( n \) large enough. This proves the proposition because \( S_M(\psi, \varphi_\infty) \geq S(\psi_1, \varphi) = +\infty. \)

Now suppose \( s(\psi_1) \leq s(\varphi) \). Then we have \( s(\psi_n) \leq \otimes^n_1 s(\psi_1) \leq \otimes^n_1 s(\varphi) = s(\varphi_n) \) for every \( n \). Define the selfadjoint operators \( H_n, n \geq 1, \) in \( \pi_\psi(\mathcal{A}_\infty)'' \) by
\[
H_n = \frac{1}{n} \left\{ \pi_\psi \left( \log \frac{d\psi_n}{d\text{Tr}_n} \right) - \pi_\psi \left( \log \frac{d\varphi_n}{d\text{Tr}_n} \right) \right\} \\
= \frac{1}{n} \pi_\psi \left( \log \frac{d\psi_n}{d\text{Tr}_n} \right) - \frac{1}{n} \sum_{i=0}^{n-1} \tilde{\gamma}_i^i \left( \pi_\psi \left( \log \frac{d\varphi}{d\text{Tr}} \right) \right).
\]

Lemma 10 and the noncommutative mean ergodic theorem imply that \( H_n \) converges in \( \bar{\psi} \)-measure as \( n \to \infty \) to
\[
h_0 = \lim_{n \to \infty} \frac{1}{n} \psi_n \left( \log \frac{d\psi_n}{d\text{Tr}_n} - \log \frac{d\varphi_n}{d\text{Tr}_n} \right) = S_M(\psi, \varphi_\infty).
\]

Let \( 0 < \delta < 1 - \varepsilon. \) For each \( n \) let \( p_n \) be the projection in \( \mathcal{A}_n \) with \( p_n \leq s(\psi_n) \) such that \( \pi_\psi(p_n) \) is the spectral projection of \( H_n \) corresponding to \( (h_0 - \delta, h_0 + \delta) \). Then there exists \( n_0 \) such that \( \psi(p_n) \geq 1 - \delta \) for all \( n \geq n_0. \) Since \( d\psi_n/d\text{Tr}_n, d\varphi_n/d\text{Tr}_n \) and \( p_n \) commute with one another, we get as (22),
\[
p_n \geq e^{-n(h_0+\delta)} p_n \left( \log \frac{d\psi_n}{d\text{Tr}_n} \right) \left( \log \frac{d\varphi_n}{d\text{Tr}_n} \right)^{-1}.
\tag{27}
\]

For each \( n \) choose a projection \( q_n \) in \( \mathcal{A} \) such that \( \psi(q_n) \geq 1 - \varepsilon \) and
\[
\varphi_n(q_n) < \exp \{ \beta_\varepsilon(\psi_n, \varphi_n) + 1 \}. \tag{28}
\]
Let $q'_n = q_n \land p_n$. Then for $n > n_0$ we get thanks to the traciality of $\psi$,

$$\psi(1 - q'_n) \leq \psi(1 - q_n) + \psi(1 - p_n) \leq \varepsilon + \delta.$$  \hfill (29)

Furthermore by (27)

$$\varphi_n(q_n) \geq \varphi_n(q'_n) = \text{Tr}_n \left( q'_n \left( p_n \frac{d\varphi_n}{d\text{Tr}_n} \right) q'_n \right)$$

$$\geq e^{-n(h_0 + \delta)} \text{Tr}_n \left( q'_n \left( p_n \frac{d\varphi_n}{d\text{Tr}_n} \right) q'_n \right)$$

$$= e^{-n(h_0 + \delta)} \text{Tr}_n \left( q'_n \frac{d\varphi_n}{d\text{Tr}_n} \right) = e^{-n(h_0 + \delta)} \varphi_n(q'_n).$$ \hfill (30)

By (28)–(30) we have for $n \geq n_0$

$$\exp\left\{ \beta \varepsilon \left( \psi_n, \varphi_n \right) + 1 \right\} \geq e^{-n(h_0 + \delta)}(1 - \varepsilon - \delta),$$

so that since $1 - \varepsilon - \delta > 0$,

$$\frac{1}{n} \left\{ \beta \varepsilon \left( \psi_n, \varphi_n \right) + 1 \right\} \geq -h_0 - \delta + \frac{1}{n} \log(1 - \varepsilon - \delta).$$

Therefore

$$\liminf_{n \to \infty} \frac{1}{n} \beta \varepsilon \left( \psi_n, \varphi_n \right) \geq -h_0 - \delta.$$  \hfill \Box

This implies (26).

Acknowledgments

The first author wishes to thank Professor Hiroshi Nagaoka who suggested to him the asymptotic theory for relative entropy. The second author is grateful to Professors L. Accardi and I. Csiszár for discussions on abelian subalgebras and I-divergency. The authors would also like to thank the referee for many valuable remarks on the first version of this paper.

References

60  Fumio Hiai and Dénes Petz


[27] D. Petz, “Characterization of the relative entropy of states of matrix algebras,” preprint.\(^2\)


**Technical Terms in Operator Algebra\(^8\)**

*operator-concavity*

A real continuous function $f$ on an interval $I$ is said to be *operator concave* if the inequality $f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda)f(B)$ in the positive semidefiniteness order holds for any selfadjoint operators $A, B$ with spectra inside $I$ and for any $0 < \lambda < 1$, where $f(A)$ and others are defined via continuous functional calculus.

*AF $C^\ast$-algebra*

A $C^\ast$-algebra $\mathcal{A}$ is said to be *AF* (approximately finite) if there is a net (usually a sequence) $\{\mathcal{A}_j\}$ of finite dimensional $*$-subalgebras of $\mathcal{A}$ such that $\bigcup_j \mathcal{A}_j$ is norm-dense in $\mathcal{A}$.

*hyperfinite von Neumann algebra*

Let $\mathcal{M}$ be a von Neumann algebra on a separable Hilbert space $\mathcal{H}$, i.e., $\mathcal{M}$ is a $*$-subalgebra (containing the identity operator) of $B(\mathcal{H})$ such that $\mathcal{M} = \mathcal{M}''$, the double commutant of $\mathcal{M}$, or equivalently $\mathcal{M}$ is closed in the

---

\(^1\)Editorial note: This paper has been published as D. Petz, “On certain properties of the relative entropy of states of operator algebras,” *Math. Z.*, 206, 351–361 (1991).


\(^8\)This part is written by F. Hiai based on editor’s special request.
weak operator topology. If there is an increasing sequence \( \{ M_n \} \) of finite dimensional \(*\)-subalgebras of \( \mathcal{M} \) such that \( (\bigcup_{n=1}^{\infty} M_n)'' = \mathcal{M} \), \( \mathcal{M} \) is said to be hyperfinite or AFD (approximately finite dimensional). A fundamental result of Connes says that \( \mathcal{M} \) is hyperfinite if and only if it is injective, i.e., there is a norm one projection from \( B(\mathcal{H}) \) onto \( \mathcal{M} \).

**GNS representation**

Let \( \mathcal{A} \) be a \( C^* \)-algebra with identity 1 and \( \varphi \) be a state on \( \mathcal{A} \). A Hilbert space \( \mathcal{H}_\varphi \) is defined as the completion of \( \mathcal{A}/\mathcal{N}_\varphi \) with an inner product \( \langle a + \mathcal{N}_\varphi, b + \mathcal{N}_\varphi \rangle = \varphi(b^*a), a, b \in \mathcal{A}, \) where \( \mathcal{N}_\varphi = \{ a \in \mathcal{A} : \varphi(a^*a) = 0 \} \). The \(*\)-representation \( \pi_\varphi : \mathcal{A} \rightarrow B(\mathcal{H}_\varphi) \) can be defined by \( \pi_\varphi(a)(b + \mathcal{N}_\varphi) = ab + \mathcal{N}_\varphi, a, b \in \mathcal{A}, \) so that \( \varphi(a) = \langle \pi_\varphi(a) \xi_\varphi, \xi_\varphi \rangle \) and \( \mathcal{H}_\varphi = \pi_\varphi(\mathcal{A})\xi_\varphi \) where \( \xi_\varphi = 1 + \mathcal{N}_\varphi \). Such a triple \( (\pi_\varphi, \mathcal{H}_\varphi, \xi_\varphi) \) is unique up to unitary equivalence, and it is called the GNS (Gelfand-Naimark-Segal) representation of \( \mathcal{A} \) associated with \( \varphi \).

**Mean ergodic theorem**

The mean ergodic theorem of von Neumann says that if \( U \) is a unitary operator on a Hilbert space \( \mathcal{H} \) and \( P \) is the orthogonal projection form \( \mathcal{H} \) onto the subspace \( \{ \xi \in \mathcal{H} : U\xi = \xi \} \), then \( n^{-1} \sum_{k=0}^{n-1} U^k \) converges to \( P \) in the strong operator topology. When \( \sigma \) is an automorphism of a \( C^* \)-algebra \( \mathcal{A} \) and \( \varphi \) is a state on \( \mathcal{A} \) with \( \varphi \circ \sigma = \varphi \), there is a unitary operator \( U_\sigma \) on \( \mathcal{H}_\varphi \) implementing \( \sigma \) so that \( \pi_\varphi(\sigma(a)) = U_\sigma \pi_\varphi(a) U_\sigma^*, a \in \mathcal{A}, \) where \( (\pi_\varphi, \mathcal{H}_\varphi, \xi_\varphi) \) is the GNS representation associated with \( \varphi \). Since \( \pi_\varphi(\sigma^k(a))\xi_\varphi = U_\sigma^k \pi_\varphi(a)\xi_\varphi \), the mean ergodic theorem can be applied to speak the convergence of \( \pi_\varphi(\sigma^k(a))\xi_\varphi \).

**Kolmogorov-Sinai entropy**

Let \((X, \mathcal{B}, \mu)\) be a probability space and \( T \) be an automorphism on \( X \) with \( \mu \circ T^{-1} = \mu \). For any finite measurable partition \( \mathcal{A} \) of \( X \), let \( H_\mu(\bigvee_{k=0}^{n-1} T^{-k} \mathcal{A}) \) denote the (Shannon) entropy of \( \bigvee_{k=0}^{n-1} T^{-k} \mathcal{A} \), the finite partition generated by \( \mathcal{A}, T^{-1} \mathcal{A}, \ldots, T^{-(n-1)} \mathcal{A}, \) with respect to \( \mu \). The limit \( h_\mu(\mathcal{A}, T) = \lim_{n \to \infty} n^{-1} H_\mu(\bigvee_{k=0}^{n-1} T^{-k} \mathcal{A}) \) exists, and the Kolmogorov-Sinai entropy \( h_\mu(T) \) is defined as the supremum of \( h_\mu(\mathcal{A}, T) \) over all finite measurable partitions \( \mathcal{A} \). If \( \mathcal{A}_0 \) is a measurable partition such that \( \bigvee_{k=-\infty}^{\infty} T^{-k} \mathcal{A}_0 = \mathcal{B} \), then \( h_\mu(T) = h_\mu(\mathcal{A}_0, T) \) (the Kolmogorov-Sinai theorem). In particular, when \( X \) is the infinite product space of \( \{1, 2, \ldots, d\} \) (the
alphabets) with the shift $T$ and $\mu$ is a shift-invariant probability measure on $X$, $h_\mu(T)$ is the mean entropy of $\mu$ or the average information contained in the input-source $\mu$ per letter.

**Shannon-McMillan-Breiman theorem**

Let $(X, \mathcal{B}, \mu, T)$ be as in the item of Kolmogorov-Sinai entropy. For each finite measurable partition $\mathcal{A}$ of $X$, let $I_\mu(\mathcal{A})$ be the information of $\mathcal{A}$ with respect to $\mu$ defined by $I_\mu(\mathcal{A}) = -\sum_{A \in \mathcal{A}} \chi_A(x) \log \mu(A)$ (so the expectation of $I_\mu(\mathcal{A})$ is the entropy $H_\mu(\mathcal{A})$). The **Shannon-McMillan-Breiman theorem** asserts that if $T$ is ergodic then $n^{-1} I_\mu(\bigvee_{k=0}^{n-1} T^{-k} \mathcal{A})$ converges to $h_\mu(\mathcal{A}, T)$ in $L^1$ and $\mu$-almost surely. This theorem played a crucial role in information theory and coding theory.
CHAPTER 4

Strong Converse Theorems in Quantum Information Theory

Hiroshi Nagaoka

Abstract. There are many situations in information theory where it is required to find the threshold value of some relevant quantity for which a kind of error probability can asymptotically be made arbitrarily small. The strong converse property means that when the quantity exceeds the threshold value the error probability inevitably goes to one asymptotically. In quantum information theory, the strong converse property is known for (classical) capacity of quantum memoryless channel ([1, 4]) and for hypothesis testing problem of two quantum i.i.d. states ([2]). In this article we provide new proofs, which are based on a simple and unified argument, with these results.

1. Key Inequality

Let \( \rho \) and \( \sigma \) be density operators on a common Hilbert space \( \mathcal{H} \). For any \( s \in [0, 1] \) and any POVM \( \{X_i\}_{i=1}^k \) it holds that \( \sum_{i=1}^k (\Tr[\rho X_i])^{1+s} (\Tr[\sigma X_i])^{-s} \leq \Tr[\rho^{1+s} \sigma^{-s}] \), which follows from a result in [3] and was used to prove the strong converse theorem for quantum hypothesis testing in [2]. In particular, for any self-adjoint operator \( A \) satisfying \( 0 \leq A \leq I \), we call a test in general, we have

\[
(\Tr[\rho A])^{1+s} (\Tr[\sigma A])^{-s} \leq \Tr[\rho^{1+s} \sigma^{-s}].
\]

This is the key inequality for the arguments below.

2. Hypothesis Testing

Given a test \( A_n \) on \( \mathcal{H}^{\otimes n} \), let \( \alpha_n(A_n) \equiv 1 - \Tr[\rho^{\otimes n} A_n] \) and \( \beta_n(A_n) \equiv \Tr[\sigma^{\otimes n} A_n] \), which are interpreted as the error probabilities of the first and second kinds, and apply the above inequality to \( \rho^{\otimes n} \), \( \sigma^{\otimes n} \) and \( A_n \) to yield

\[
\frac{1+s}{n} \log(1-\alpha_n(A_n)) - \frac{s}{n} \log \beta_n(A_n) \leq \psi(s),
\]

where \( \psi(s) \equiv \log \Tr[\rho^{1+s} \sigma^{-s}] \). From \( \lim_{s \downarrow 0} \psi(s)/s = \Tr[\rho(\log \rho - \log \sigma)] \equiv D(\rho || \sigma) \), we see that if \( r \equiv -\limsup_{n \to \infty} \frac{1}{n} \log \beta_n(A_n) > D(\rho || \sigma) \) then \( \alpha_n(A_n) \) necessarily goes to one with exponential rate

\[
\limsup_{n \to \infty} \frac{1}{n} \log(1-\alpha_n(A_n)) \leq -\max_{0 \leq s \leq 1} \left\{ \frac{s}{1+s} r - \frac{\psi(s)}{1+s} \right\}.
\]
3. Channel Coding

Let \( \Delta \) be a compact subset of \( S(\mathcal{H}) \) (the set of density operators on \( \mathcal{H} \)), which is supposed to be the set of possible received states of a quantum communication channel. For \( n, M \in \mathbb{N} \) and \( \varepsilon > 0 \), an \((n,M,\varepsilon)\)-code consists of a family of product states \( \rho_{i}^{(n)} = \rho_{i,1} \otimes \cdots \otimes \rho_{i,n} \) with \( \rho_{i,j} \in \Delta, \ i = 1, \ldots, M \), and a POVM \( X^{(n)} = \{X_{i}^{(n)}\}_{i=1}^{M} \) on \( \mathcal{H}^{\otimes n} \) satisfying \( \frac{1}{M} \sum_{i} \text{Tr}[\rho_{i}^{(n)} X_{i}^{(n)}] \geq 1 - \varepsilon \). Setting \( \rho, \sigma \) and \( A \) in (1) to be the block diagonal matrices \( \frac{1}{M} \text{diag}\{\rho_{1}^{(n)}, \ldots, \rho_{M}^{(n)}\}, \frac{1}{M} \text{diag}\{\sigma^{\otimes n}, \ldots, \sigma^{\otimes n}\} \) and \( \text{diag}\{X_{1}^{(n)}, \ldots, X_{M}^{(n)}\} \), where \( \sigma \) is an arbitrary state in \( S(\mathcal{H}) \), we have

\[
(1 - \varepsilon)^{1+s} M^n \leq \frac{1}{M} \sum_{i=1}^{M} \prod_{j=1}^{n} \text{Tr}[\rho_{i,j}^{1+s} \sigma^{-s}] \leq \max_{\rho \in \Delta} (\text{Tr}[\rho^{1+s} \sigma^{-s}])^n.
\]

It follows that if a sequence of \((n,M_n,\varepsilon_n)\)-codes for \( n = 1, 2, \ldots \) satisfies

\[
R \overset{\text{def}}{=} \lim_{n \to \infty} \frac{1}{n} \log M_n > \min_{\sigma \in S(\mathcal{H})} \max_{\rho \in \Delta} D(\rho||\sigma) = \max_{\lambda,\{\rho_{i}\} \subseteq \Delta} \sum_{i} \lambda_{i} D(\rho_{i}|| \sum_{j} \lambda_{j} \rho_{j}),
\]

then \( \varepsilon_n \) necessarily goes to one with exponential rate

\[
\lim_{n \to \infty} \frac{1}{n} \log(1 - \varepsilon_n) \leq - \max_{0 \leq s \leq 1} \left\{ \frac{s}{1 + s} R - \min_{\sigma \in S(\mathcal{H})} \max_{\rho \in \Delta} \frac{1}{1 + s} \log \text{Tr}[\rho^{1+s} \sigma^{-s}] \right\}.
\]

References

CHAPTER 5
Asymptotics of Quantum Relative Entropy from a Representation Theoretical Viewpoint

Masahito Hayashi

Abstract. In this paper it is proved that the quantum relative entropy $D(\rho \| \sigma)$ can be asymptotically attained by the relative entropy of probabilities given by a certain sequence of positive-operator-valued measurements (POVMs). The sequence of POVMs depends on $\sigma$, but is independent of the choice of $\rho$.

1. Introduction

In classical statistical theory, the relative entropy $D(p \| q)$ is an information quantity which represents the statistical efficiency in distinguishing a probability measure $p$ from another probability measure $q$. In the quantum mechanical case, states are described by density operators on a Hilbert space $\mathcal{H}$, which represents a physical system of interest. We can distinguish quantum states by data given through quantum measurements. A quantum measurement is described by a positive-operator-valued measure (POVM) $M = \{M_i\}_{i \in I}$, which is a partition of the unit into positive operators. A POVM $M = \{M_i\}_{i \in I}$ satisfying $M_i^2 = M_i$ for any index $i$ is called a projection-valued measure (PVM); this plays an important role in this paper. When the quantum measurement corresponding to a POVM $M$ is made on the system in a state $\rho$, the data obey the probability distribution $P^M_\rho(i) := \text{Tr} M_i \rho$.

The quantum relative entropy $D(\rho \| \sigma) := \text{Tr} \rho (\log \rho - \log \sigma)$ is known as a quantum analogue of the relative entropy. However, the information quantity, which is directly linked to statistical significance, is not the quantum relative entropy $D(\rho \| \sigma)$, but the relative entropy $D^M(\rho \| \sigma) := D(P^M_\rho \| P^M_\sigma)$. Concerning the relation between the two quantities $D(\rho \| \sigma)$ and $D^M(\rho \| \sigma)$, we have the following inequality from the monotonicity of the quantum relative entropy [5, 9]

$$D^M(\rho \| \sigma) \leq D(\rho \| \sigma)$$

(1)

where the equality holds for some $M$ if and only if $\rho \sigma = \sigma \rho$ (see Petz [8], proposition 1.16 in Ohya-Petz [6] and Fujiwara-Nagaoka [1]). As for the inequality (1), Hiai-Petz [4] proved that even if the states $\rho$ and $\sigma$ are not commutative with one another, the equality is attained in an asymptotic setting as described below. First, we introduce the quantum i.i.d. condition in order to treat an asymptotic setting. Suppose that $n$ independent physical systems are given in the same state $\rho$, then the quantum state of the composite system is described by $\rho^{\otimes n}$, defined by

$$
\rho^{\otimes n} := \rho \otimes \cdots \otimes \rho \text{ on } \mathcal{H}^{\otimes n}
$$

where the tensored space $\mathcal{H}^{\otimes n}$ is defined by

$$
\mathcal{H}^{\otimes n} := \mathcal{H} \otimes \cdots \otimes \mathcal{H}.
$$

We call this condition the quantum i.i.d. condition, which is a quantum analogue of the independent-identical-distribution condition. Under the quantum i.i.d. condition, the equation

$$
D(\rho^{\otimes n} \| \sigma^{\otimes n}) = nD(\rho \| \sigma)
$$

holds. Hiai-Petz [4] proved the following theorem. (For the infinite-dimensional case, see Petz [7].)

**Theorem 1:** Let $k$ be the dimension of $\mathcal{H}$ and let $\sigma$ and $\rho_n$ be states on $\mathcal{H}$ and $\mathcal{H}^{\otimes n}$, respectively. Then there exists a POVM $M^n$ on tensored space $\mathcal{H}^{\otimes n}$ such that

$$
\frac{1}{n} D(\rho_n \| \sigma^{\otimes n}) - \frac{(k - 1) \log(n + 1)}{n} \leq \frac{1}{n} D^{M^n}(\rho_n \| \sigma^{\otimes n}) \leq \frac{1}{n} D(\rho_n \| \sigma^{\otimes n}). \tag{2}
$$

When the limit $\lim_{n \to \infty} \frac{1}{n} D(\rho_n \| \sigma^{\otimes n})$ converges, the equation

$$
\lim_{n \to \infty} \frac{1}{n} D^{M^n}(\rho_n \| \sigma^{\otimes n}) = \lim_{n \to \infty} \frac{1}{n} D(\rho_n \| \sigma^{\otimes n})
$$

holds.

Theorem 1 tells us that there exists a sequence $\{M_n\}$ satisfying (2) which may depend on both $\{\rho_n\}$ and $\sigma$. In this paper, using a representation theoretical argument on the representation of $SL(\mathcal{H})$ on $\mathcal{H}^{\otimes n}$, we prove that there exists a sequence $\{M_n\}$ satisfying (2) which depends only on $\sigma$. 

Asymptotics of Quantum Relative Entropy
when the sequence \( \{ \rho_n \} \) satisfies the quantum i.i.d. condition; i.e., the state \( \rho_n \) is \( \rho^\otimes n \). The following is the main theorem.

**Theorem 2:** Let \( k \) be the dimension of \( \mathcal{H} \) and let \( \sigma \) be a state on \( \mathcal{H} \). Then there exists a POVM \( M^n \) on the tensored space \( \mathcal{H}^\otimes n \) which satisfies

\[
D(\rho \| \sigma) - \frac{(k - 1) \log(n + 1)}{n} \leq \frac{1}{n} D^{M^n}(\rho^\otimes n \| \sigma^\otimes n) \leq D(\rho \| \sigma) \quad \forall \rho.
\]

In section 2, we prove theorem 2 and discuss the difference between our proof and the proof of theorem 1 given by Hiai-Petz [4]. In section 3, we explain some results in representation theory which are necessary for the proof of theorem 2. In the end of section 3, we construct the POVM \( M^n \) satisfying (3). In section 4, we extend theorem 2 to the infinite-dimensional case.

If we perform the POVM \( M^n \) satisfying (3), we can attain the quantum relative entropy \( D(\rho \| \sigma) \) w.r.t. the rate of the second error probability in the quantum hypothesis testing: the null hypothesis is \( \rho^\otimes n \) and the alternative is \( \sigma^\otimes n \). Theorem 2 claims that the POVM \( M^n \) is independent of the alternative \( \rho \). Therefore, the POVM \( M^n \) is useful for the quantum hypothesis testing, in which the alternative hypothesis consists of plural tensored states. In addition, an application of theorem 2 to the quantum estimation will be discussed in another paper [3] in preparation.

2. Proof of the Main Theorem

In this section, we will prove the main theorem after some discussions about PVMs and the quantum relative entropy in the non-asymptotic setting. We make some definitions for this purpose. A state \( \rho \) is called *commutative* with a PVM \( E(= \{ E_i \}) \) on \( \mathcal{H} \) if \( \rho E_i = E_i \rho \) for any index \( i \). For PVMs \( E(= \{ E_i \}_{i \in I}), F(= \{ F_j \}_{j \in J}) \), the notation \( E \leq F \) means that for any index \( i \in I \) there exists a subset \((F/E)_i\) of the index set \( J \) such that \( E_i = \sum_{j \in (F/E)_i} F_j \). For a state \( \rho \), we denote \( E(\rho) \) by the spectral measure of \( \rho \), which can be regarded as a PVM. The map \( \mathcal{E}_E \) with respect to a PVM \( E \) is defined as:

\[
\mathcal{E}_E : \rho \mapsto \sum_i E_i \rho E_i
\]

which is an affine map from the set of states to itself. Note that the state \( \mathcal{E}_E(\rho) \) is commutative with a PVM \( E \). If a PVM \( F = \{ F_j \} \) is commutative with a PVM \( E = \{ E_i \} \), then we can define the PVM \( F \times E = \{ F_j E_i \} \), which satisfies that \( F \times E \geq E \) and \( F \times E \geq F \).
**Theorem 3:** Let $E$ be a PVM such that $w(E) := \sup_i \dim E_i < \infty$. If states $\sigma, \rho$ are commutative with the PVM $E$ and a PVM $F$ satisfies $E \leq F, E(\sigma) \leq F$, then we have

$$D(\rho\|\sigma) - \log w(E) \leq D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) \leq D(\rho\|\sigma).$$

(5)

This theorem follows from lemma 5 and lemma 6 below. Using theorem 3 and the following lemma, we will prove the main theorem.

**Lemma 4:** There exists a PVM $E^n$ on $\mathcal{H}^\otimes n$ which is commutative with $\rho^\otimes n$ for any $\rho$ and satisfies the relation $w(E^n) \leq (n + 1)^{(k-1)}$.

Lemma 4 is proved in the next section from a representation theoretical viewpoint. Now, let $E^n$ be a PVM satisfying the condition given in lemma 4. Then there exists a PVM $F^n$ such that $F^n \geq E^n \times E(\sigma^\otimes n)$ and that $w(F^n) = 1$. Using theorem 3, we have the following.

$$D(\rho\|\sigma) - \frac{(k-1)\log(n+1)}{n} \leq \frac{1}{n}D(\mathcal{E}_{F^n}(\rho^\otimes n)\|\mathcal{E}_{F^n}(\sigma^\otimes n)) \leq D(\rho\|\sigma) \quad \forall \rho. \tag{6}$$

Since the condition $w(F^n) = 1$ implies $D(\mathcal{E}_{F^n}(\rho^\otimes n)\|\mathcal{E}_{F^n}(\sigma^\otimes n)) = D^{F^n}(\rho^\otimes n\|\sigma^\otimes n)$, we obtain theorem 2.

Let us compare the above argument with that of Hiai-Petz [4]. We first note that the inequality $D^F(\rho\|\sigma) \leq D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma))$ holds for any PVM, but the equality, in general, does not hold unless $w(F) = 1$. Instead of (6), Hiai-Petz proved the following:

$$\frac{1}{n}D(\rho_n\|\sigma^\otimes n) - \frac{(k-1)\log(n+1)}{n} \leq \frac{1}{n}D(\mathcal{E}_{E(\sigma^\otimes n)}(\rho^\otimes n)\|\mathcal{E}_{E(\sigma^\otimes n)}(\sigma^\otimes n))$$

$$\leq \frac{1}{n}D(\rho_n\|\sigma^\otimes n). \tag{7}$$

In the case when $\rho_n = \rho^\otimes n$, this is the same as (6) except that $E(\sigma^\otimes n)$ is substituted for $F^n$. Since $w(E(\sigma^\otimes n)) = 1$ does not hold, however, $D(\mathcal{E}_{E(\sigma^\otimes n)}(\rho^\otimes n)\|\mathcal{E}_{E(\sigma^\otimes n)}(\sigma^\otimes n))$ cannot be replaced with $D^{E(\sigma^\otimes n)}(\rho^\otimes n\|\sigma^\otimes n)$ here. In other words, even though the PVM $E(\sigma^\otimes n)$ does not depend on a state $\rho$, the inequality (7) does not imply the existence of a PVM $M^n$ depending only on $\sigma$ and satisfying (3) for all $\rho$. Indeed, the PVM $M^n$ which was shown to satisfy (2) in [4] is not $E(\sigma^\otimes n)$, but $E(\sigma^\otimes n) \times E(\mathcal{E}_{E(\sigma^\otimes n)}(\rho_n))$, which depends on $\rho_n$ in general. Therefore, the discussion in Hiai-Petz [4] does not imply theorem 2.

Now, we prove two lemmas used in a proof of theorem 3. The following lemma 5 is the same as lemma 3.1 in Hiai-Petz [4] and theorem 1.13 in
Ohya-Petz [6]. However, lemma 5 is proved in the following because it plays a particularly important role in our proof of the main theorem.

**Lemma 5:** Let $\rho, \sigma$ be states. If a PVM $F$ satisfies $E(\sigma) \leq F$, then

$$D(\rho\|\sigma) = D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) + D(\rho\|\mathcal{E}_F(\rho)). \quad (8)$$

**Proof:** As $E(\sigma) \leq F$ and $F$ is commutative with $\sigma$, we have $\text{Tr} \mathcal{E}_F(\rho) \log \mathcal{E}_F(\sigma) = \text{Tr} \rho \log \sigma$. Since $\rho$ is commutative with $\log \rho$, we have $\text{Tr} \mathcal{E}_F(\rho) \log \rho = \text{Tr} \rho \log \rho$. Therefore, we get the following:

$$D(\mathcal{E}_F(\rho)\|\mathcal{E}_F(\sigma)) - D(\rho\|\sigma)
= \text{Tr} \mathcal{E}_F(\rho) (\log \mathcal{E}_F(\rho) - \log \mathcal{E}_F(\sigma)) - \text{Tr} \rho (\log \rho - \log \sigma)
= \text{Tr} \mathcal{E}_F(\rho) (\log \mathcal{E}_F(\rho) - \log \rho).$$

This proves (8).

**Lemma 6:** Let $E, F$ be PVMs such that $E \leq F$. If a state $\rho$ is commutative with $E$, then we have

$$D(\rho\|\mathcal{E}_F(\rho)) \leq \log w(E). \quad (9)$$

**Proof:** Let $a_i := \text{Tr} E_i \rho E_i, \rho_i := \frac{1}{a_i} E_i \rho E_i$. Then we have $\rho = \sum_i a_i \rho_i$, $\mathcal{E}_F(\rho) = \sum_i a_i \mathcal{E}_F(\rho_i), \sum_i a_i = 1$. Therefore,

$$D(\rho\|\mathcal{E}_F(\rho)) = \sum_i \text{Tr} E_i \rho (\log \rho - \log \mathcal{E}_F(\rho))
= \sum_i \text{Tr} E_i \rho E_i (E_i \log \rho E_i - E_i \log \mathcal{E}_F(\rho) E_i)
= \sum_i a_i D(\rho_i\|\mathcal{E}_F(\rho_i)) \leq \sup_i D(\rho_i\|\mathcal{E}_F(\rho_i))
= \sup_i (\text{Tr} \rho_i \log \rho_i - \text{Tr} \mathcal{E}_F(\rho_i) \log \mathcal{E}_F(\rho_i))
\leq - \sup_i \text{Tr} \mathcal{E}_F(\rho_i) \log \mathcal{E}_F(\rho_i) \leq \sup_i \log \dim E_i = \log w(E).$$

Thus, we obtain inequality (9).

It is interesting to compare (9) with lemma 3.2 in Hiai-Petz [4], which was used to show (7); i.e., $D(\rho\|\mathcal{E}_F(\rho)) \leq \log h(F)$ where $h(F)$ denotes the number of indices $i \in I$ for the PVM $F = \{F_i\}_{i \in I}$. 
3. Quantum I.I.D. Condition from Group Theoretical Viewpoint

In this section, we discuss the quantum i.i.d. condition from a group theoretical viewpoint for the purpose of lemma 4. In section 3.1, we consider the relation between irreducible representations and PVMs. In section 3.2, we discuss the quantum i.i.d. condition and PVMs from a theoretical viewpoint.

3.1. Group representation and its irreducible decomposition

Let $V$ be a finite-dimensional vector space over the complex numbers $\mathbb{C}$. A map $\pi$ from a group $G$ to the generalized linear group of a vector space $V$ is called a representation on $V$ if the map $\pi$ is homomorphism, i.e., $\pi(g_1)\pi(g_2) = \pi(g_1g_2)$, $\forall g_1, g_2 \in G$. A subspace $W$ of $V$ is called invariant with respect to a representation $\pi$ if the vector $\pi(g)w$ belongs to the subspace $W$ for any vector $w \in W$ and any element $g \in G$. A representation $\pi$ is called irreducible if there is no proper nonzero invariant subspace of $V$ with respect to $\pi$. Let $\pi_1$ and $\pi_2$ be representations of a group $G$ on $V_1$ and $V_2$, respectively. The tensored representation $\pi_1 \otimes \pi_2$ of $G$ on $V_1 \otimes V_2$ is defined as $(\pi_1 \otimes \pi_2)(g) = \pi_1(g) \otimes \pi_2(g)$, and the direct sum representation $\pi_1 \oplus \pi_2$ of $G$ on $V_1 \oplus V_2$ is also defined as $(\pi_1 \oplus \pi_2)(g) = \pi_1(g) \oplus \pi_2(g)$.

In the following, we treat a representation $\pi$ of a group $G$ on a finite-dimensional Hilbert space $\mathcal{H}$. The following facts are crucial in the later arguments. There exists an irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ such that the irreducible components are orthogonal to one another if for any element $g \in G$ there exists an element $g^* \in G$ such that $\pi(g)^* = \pi(g^*)$, where $\pi(g)^*$ denotes the adjoint of the linear map $\pi(g)$. We can regard the irreducible decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_l$ as the PVM $\{P_{\mathcal{H}_i}\}_{i=1}^l$, where $P_{\mathcal{H}_i}$ denotes the projection to $\mathcal{H}_i$. If two representations $\pi_1, \pi_2$ satisfy the preceding condition, then the tensored representation $\pi_1 \otimes \pi_2$ also satisfies it. Note that, in general, an irreducible decomposition of a representation satisfying the preceding condition is not unique. In other words, we cannot uniquely define the PVM from such a representation.

3.2. Relation between the tensored representation and PVMs

Let the dimension of the Hilbert space $\mathcal{H}$ be $k$. Concerning the natural representation $\pi_{SL(\mathcal{H})}$ of the special linear group $SL(\mathcal{H})$ on $\mathcal{H}$, we consider...
its $n$th tensored representation $\pi_{\otimes n}^{\otimes n}_{SL(H)} := \pi_{SL(H)} \otimes \cdots \otimes \pi_{SL(H)}$ on the tensored space $H^{\otimes n}$. For any element $g \in SL(H)$, the relation $\pi_{SL(H)}(g^*) = \pi_{SL(H)}(g^*)$ holds, where the element $g^* \in SL(H)$ denotes the adjoint matrix of the matrix $g$. Consequently, there exists an irreducible decomposition of $\pi_{\otimes n}^{\otimes n}_{SL(H)}$ regarded as a PVM and we denote the set of such PVMs by $\text{Ir}_{\otimes n}$.

From the Weyl dimension formula ((7.1.8) or (7.1.17) in Goodman-Walch [2]), the $n$th symmetric tensored space is the maximum-dimensional space in the irreducible subspaces with respect to the $n$th tensored representation $\pi_{\otimes n}^{\otimes n}_{SL(H)}$. Its dimension is equal to the repeated combination $k\cdot \pi_{n}$ evaluated by $kH_n$.

**Lemma 7:** A PVM $E^n \in \text{Ir}_{\otimes n}$ is commutative with the $n$th tensored state $\rho_{\otimes n}$ of any state $\rho$ on $H$.

**Proof:** If $\det \rho \neq 0$, then this lemma is trivial from the fact that $\det(\rho)^{-1} \rho \in SL(H)$. If $\det \rho = 0$, there exists a sequence $\{\rho_i\}_{i=1}^{\infty}$ such that $\det \rho_i \neq 0$ and $\rho_i \to \rho$ as $i \to \infty$. We have $\rho_i^{\otimes n} \to \rho^{\otimes n}$ as $i \to \infty$. Because a PVM $E^n \in \text{Ir}_{\otimes n}$ is commutative with $\rho_i^{\otimes n}$, it is, also, commutative with $\rho^{\otimes n}$. \[ \blacksquare \]

Now, we can see that lemma 4 follows from lemma 7 and (10).

From the above discussion, if the PVM $M^n$ satisfies $w(M^n) = 1$ and $E(\sigma^{\otimes n}) \times E^n \leq M^n$ for some $E^n \in \text{Ir}_{\otimes n}$, the inequality (3) holds. In many cases, the relation $w(E(\sigma^{\otimes n}) \times E^n) = 1$ holds. Therefore, in these cases, the PVM $E(\sigma^{\otimes n}) \times E^n$ satisfies (3). For example, if the eigenvalues of $\sigma$ are rationally independent of each other, the relation holds for arbitrary $n$. Also, in the spin-1/2 system except $\sigma = \frac{1}{2} \text{Id}$, the relation holds. In this case, the POVM $E(\sigma^{\otimes n}) \times E^n$ can be regarded as a simultaneous measurement of the total momentum and the momentum of the direction specified by $\sigma$.

4. Infinite-Dimensional Case

Next, we prove an infinite-dimensional version of theorem 2. Let $\mathcal{B}(H)$ be the set of bounded operators on $H$, and $\mathcal{B}(H)^{\otimes n}$ be $\mathcal{B}(H) \otimes \cdots \otimes \mathcal{B}(H)$. According to [6], from the separability of $H$, there exists a finite-dimensional approximation of $H$, i.e., a sequence $\{\alpha_n : \mathcal{B}(H_n) \to \mathcal{B}(H)\}$ of unit-
preserving completely positive maps such that $\mathcal{H}_n$ is finite-dimensional and
\begin{equation}
\lim_{n \to \infty} D(\alpha_n^*(\rho)\|\alpha_n^*(\sigma)) = D(\rho\|\sigma) \tag{11}
\end{equation}
for any states $\rho, \sigma$ on $\mathcal{H}$ such that $\mu \sigma \leq \rho \leq \lambda \sigma$ for some positive real numbers $\mu, \lambda$, where $\alpha_n^*$ is the adjoint of $\alpha_n$. From (3) and (11), for any positive integer $n$ there exists a pair $(l_n, M^{n'}_n := \{M^{n'}_i\})$ of an integer and a POVM on $\mathcal{H}_n^{\otimes l_n}$ such that
\begin{equation}
D(\alpha_n^*(\rho)\|\alpha_n^*(\sigma)) - \frac{D^{M^{n'}}(\alpha_n^*(\rho)^{\otimes l_n} || (\alpha_n^*(\sigma)^{\otimes l_n}))}{l_n} < \frac{1}{n}. \tag{12}
\end{equation}
The completely positive map $\alpha^{\otimes l}_n$ from $\mathcal{B}(\mathcal{H}_n)^{\otimes l}$ to $\mathcal{B}(\mathcal{H})^{\otimes l}$ is defined as $\alpha^{\otimes l}_n(A_1 \otimes A_2 \otimes \cdots \otimes A_l) = \alpha_n(A_1) \otimes \alpha_n(A_2) \otimes \cdots \otimes \alpha_n(A_l)$ for $\forall A_i \in \mathcal{B}(\mathcal{H})$. We have $(\alpha^{\otimes l}_n)^* (\rho^{\otimes l}) = \alpha^*_n(\rho)^{\otimes l}$. Let $M_n := \{\alpha_n^{\otimes l_n}(M^{n'}_i)\}$, then from (11), (12) we obtain
\begin{equation}
\frac{D^{M^n}(\rho^{\otimes l_n} || \sigma^{\otimes l_n})}{l_n} \leq \frac{D^{M^{n'}}(\alpha_n^*(\rho)^{\otimes l_n} || (\alpha_n^*(\sigma)^{\otimes l_n})^{\otimes l_n})}{l_n} < \frac{1}{n} \to D(\rho\|\sigma) \text{ as } n \to \infty.
\end{equation}
Therefore, we obtain an infinite-dimensional version of theorem 2. Note that such a POVM $M^n$ is independent of $\rho$.

5. Conclusions
It is proved that the quantum relative entropy $D(\rho\|\sigma)$ is attained by the sequence of the relative entropies given by a certain sequence of PVMs which is independent of $\rho$. This formula is closely related to the quantum asymptotic detection. The physical realization of the sequence of measurements corresponding to PVMs satisfying (3) are left for future study. In the spin-1/2 system, it follows from the representation theoretical viewpoint in section 3 that it is enough to simultaneously measure the total momentum and the momentum of the direction specified by $\sigma$.

Acknowledgments
The author wishes to thank Professor H. Nagaoka for encouragement and useful advice to complete this manuscript. He is also grateful to thank Professor D. Petz for essential suggestions about the extension in the infinite-
dimensional case. He is indebted to Professor A. Fujiwara and Dr. M. Ishii for several discussions on this topic.

References


*Editorial note: The paper [3] has been published, and is reprinted in Chap. 28 of this book.*
CHAPTER 6

Quantum Birthday Problems:
Geometrical Aspects of Quantum Random Coding

Akio Fujiwara

Abstract. This correspondence explores asymptotics of randomly generated vectors on extended Hilbert spaces. In particular, we are interested to know “orthogonal” these vectors are. We investigate two types of asymptotic orthogonality, the weak orthogonality and the strong orthogonality, that are regarded as quantum analogs of the classical birthday problem and its variant. As regards the weak orthogonality, a new characterization of the von Neumann entropy is derived, and a mechanism behind the noiseless quantum channel coding theorem is clarified. As regards the strong orthogonality, on the other hand, a characterization of the quantum Rényi entropy of degree 2 is derived.

Index Terms: Asymptotic orthogonality, birthday problem, quantum channel coding, quantum information theory, random vector, Rényi entropy, von Neumann entropy.

1. Introduction

Let $\mathcal{H}$ be a Hilbert space and let $p$ be a probability measure on $\mathcal{H}$ the support of which is a countable set $\mathcal{X}$ of unit vectors. We assume that $|\mathcal{X}| \geq 2$ and that vectors in $\mathcal{X}$ are not parallel to each other. Associated with the probability measure $p$ is the density operator

$$\rho := \sum_{\phi \in \mathcal{X}} p(\phi) |\phi\rangle \langle \phi|,$$

where, and in what follows, we use the Dirac notation. Let $\{X_k(i)\}_{k_i}$ be $\mathcal{X}$-valued random variables independent and identically distributed (i.i.d.) with respect to $p$, and let $\{L_n\}_n$ be an increasing sequence of natural numbers. For each $n \in \mathbb{N}$, we define $L_n$ random vectors $\{\Psi^{(n)}(i)\}_{1 \leq i \leq L_n}$ on $\mathcal{H}^\otimes n$ by

$$\Psi^{(n)}(i) := X_1(i) \otimes X_2(i) \otimes \cdots \otimes X_n(i), \quad (i = 1, \ldots, L_n).$$

We denote the inner product of the $i$th and $j$th vectors by
\[ g_{ij}^{(n)} := \langle \Psi^{(n)}(i) | \Psi^{(n)}(j) \rangle = \prod_{k=1}^{n} \langle X_k(i) | X_k(j) \rangle. \] (2)

Note that $g_{ii}^{(n)} = 1$ for all $n$ and $1 \leq i \leq L_n$, and that, for fixed $i$ and $j$ ($i \neq j$), $g_{ij}^{(n)}$ converges to 0 almost surely as $n \to \infty$. Thus it is natural to inquire how “orthogonal” those random vectors are. For later convenience, we denote the ordered list of the $L_n$ random vectors by $C^{(n)}$.

To motivate our problem, let us consider the special case when $X$ forms an orthonormal system. In this case, $g_{ij}^{(n)} = 0$ if $X_k(i) \neq X_k(j)$ for some $k$, and $g_{ij}^{(n)} = 1$ otherwise. Put differently, the Gram matrix
\[
G^{(n)} := \begin{bmatrix}
  g_{11}^{(n)} & \cdots & g_{1L_n}^{(n)} \\
  \vdots & & \vdots \\
  g_{L_n1}^{(n)} & \cdots & g_{L_nL_n}^{(n)}
\end{bmatrix}
\]
gives a yes/no table indicating whether the $i$th $n$-tuple $(X_1(i), ..., X_n(i))$ and the $j$th $n$-tuple $(X_1(j), ..., X_n(j))$ are identical ($g_{ij}^{(n)} = 1$) or not ($g_{ij}^{(n)} = 0$). As a consequence, orthogonality problems for the random vectors are reduced to combinatorial ones when $X$ is orthonormal.

In his paper [8], Rényi posed several combinatorial problems that can be regarded as asymptotic versions of the classical birthday problem (cf. [1]) and its variants, and characterized classical entropies. From among these, let us recast two problems in terms of orthogonalities of random vectors. Let $X$ be orthonormal. We say that $C^{(n)}$ satisfies a **weak orthogonality** condition with respect to the $i$th vector if the event
\[ E_i^{(n)} := \{ \text{the } i\text{th vector } \Psi^{(n)}(i) \text{ is orthogonal to the other vectors in } C^{(n)} \} \]
occurs, and that $C^{(n)}$ satisfies a **strong orthogonality** condition if the event
\[ F^{(n)} := \{ \text{the vectors in } C^{(n)} \text{ are mutually orthogonal} \} \]
occurs. We are interested to know how fast can $L_n$ be increased under the condition that the probability $P(E_i^{(n)})$ for some (then any) $i$, or $P(F^{(n)})$, tends to 1 as $n \to \infty$. Let $C_w(p)$ [resp., $C_s(p)$] be the supremum of
\[
\limsup_{n \to \infty} \log L_n/n \text{ over all sequences } \{L_n\}_n \text{ that satisfy } P(E_i^{(n)}) \to 1 \\
\text{[resp., } P(F^{(n)}) \to 1].
\]
We may call $C_w(p)$ [resp., $C_s(p)$] the **weak** [resp., **strong**] orthogonality capacity of the probability measure $p$. Since we are now dealing with a probability measure $p$ that has an orthonormal support $X$, the problems are essentially combinatorial, and it is not too difficult to
show that $C_w(p) = H(p)$ and $C_s(p) = H_2(p)/2$, where $H(p)$ and $H_2(p)$ are the Shannon entropy and the Rényi entropy [8] of degree 2 with respect to the probability measure $p$.

Let us now return to the general case when $X$ is not necessarily orthonormal. Although the vectors in $C^{(n)}$ may not be strictly orthogonal in this case, it would be quite possible that they are “almost” orthogonal for sufficiently large $n$. It is, therefore, expected that quantum entropies might be characterized via asymptotic properties of the set $C^{(n)}$ of random vectors (as Rényi did for classical entropies via combinatorics). The purpose of this paper is to extend the notions of weak and strong orthogonality of random vectors to a general probability measure $p$, and to determine the corresponding capacities. In fact, with proper definitions of “asymptotic” orthogonalities, it is shown in Theorems 1 and 4 that the weak orthogonality capacity $C_w(p)$ is given by the von Neumann entropy

$$H(\rho) := -\text{Tr} \rho \log \rho,$$

and the strong orthogonality capacity $C_s(p)$ is given by half the quantum Rényi entropy of degree 2

$$H_2(\rho) := -\log \text{Tr} \rho^2,$$

where the probability measure $p$ and the density operator $\rho$ are connected by (1). Since these results obviously generalize the above mentioned classical characterizations by Rényi, our problems may be called quantum birthday problems. It should be emphasized that each capacity depends only on the density operator $\rho$, so that, for probability measures $p$ and $q$ which give the same density operator, $C_w(p) = C_w(q)$ and $C_s(p) = C_s(q)$ hold.

The orthogonality of vectors in $C^{(n)}$ is closely related to their “distinguishability” in quantum measurement theory. Let a physical system of interest be represented by the Hilbert space $H^{\otimes n}$, and let a unit vector in $H^{\otimes n}$ correspond to a quantum pure state. When $X$ is orthonormal, any two vectors in $C^{(n)}$ are either orthogonal or identical, so that there is a quantum-mechanical measurement on $H^{\otimes n}$ that distinguishes distinct vectors in $C^{(n)}$ with probability one. In this sense, strict orthogonality implies strict distinguishability by a certain measurement. As a matter of fact, this corresponds to the classical situation: unlimited distinguishability for distinct objects is precisely the central dogma of the classical theory, and one can restore Rényi’s original problems by replacing the word(s) “orthogonal (to)” with “distinct (from)” in the above definitions of the events $E_i^{(n)}$ and $F^{(n)}$. When $X$ is not orthonormal, on the other hand, distinct vectors in
$C^{(n)}$ are not necessarily orthogonal, so that they are not always strictly distinguishable in the sense of quantum mechanics. We therefore have to deal with, so to say, “asymptotic distinguishability” of random vectors. In fact, we will clarify a close connection between asymptotic orthogonality and the noiseless quantum channel coding problem.

2. Weak Orthogonality

Let us start with a preliminary consideration as to what is the proper extension of the notion of weak orthogonality to the case when $X$ is not necessarily orthonormal. If the $i$th vector in $C^{(n)}$ is “almost” orthogonal to the other vectors in $C^{(n)}$, then the inner products $g_{ij}^{(n)}$, $(j = 1, \ldots, L_n, j \neq i)$, must all be sufficiently small. Thus the proper extension of weak orthogonality might be such that the random variables $g_{ij}^{(n)}$ converge to 0 simultaneously for all $j(\neq i)$ as $n \to \infty$ in a certain mode of convergence. For example, the condition that

$$Y_i^{(n)} := \sum_{j(\neq i)}^{L_n} |g_{ij}^{(n)}|^2 \to 0 \text{ in probability} \quad (3)$$

might be a candidate. However, in anticipation of a characterization of the von Neumann entropy as well as a relationship with the quantum channel coding problem, we adopt a slightly different approach. (In fact, it is shown in Appendix A that the condition (3) does not characterize the von Neumann entropy.)

For each $n$, let $\mathcal{L}^{(n)}$ be a subspace of $\mathcal{H}^{\otimes n}$ and let $\Pi_{\mathcal{L}^{(n)}}$ be the projection operator onto the subspace $\mathcal{L}^{(n)}$. Given a pair $(C^{(n)}, \mathcal{L}^{(n)})$, let us denote the inner product of the projected $i$th and $j$th vectors by

$$\hat{g}_{ij}^{(n)} := \langle \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(i) \vert \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(j) \rangle,$$

and define random variables

$$\hat{Y}_i^{(n)} := \sum_{j(\neq i)}^{L_n} |\hat{g}_{ij}^{(n)}|^2.$$

We say that a sequence $\{C^{(n)}\}_n$ satisfies the asymptotic weak orthogonality condition if there is a sequence $\{\mathcal{L}^{(n)}\}_n$ of subspaces such that the following conditions are satisfied for all $i$:

(i) $\hat{g}_{ii}^{(n)} \to 1$ in probability,

(ii) $\hat{Y}_i^{(n)} \to 0$ in probability.
Some remarks are in order. Condition (i) implies that, for sufficiently large \( n \), the \( i \)th vector \( \Psi^{(n)}(i) \) will be almost parallel to \( \mathcal{L}^{(n)} \), so that the projected \( i \)th vector \( \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(i) \) be almost identical to the original vector \( \Psi^{(n)}(i) \). Condition (ii) implies that, for sufficiently large \( n \), all the vectors \( \Psi^{(n)}(j), (j \neq i) \), will be simultaneously almost orthogonal to the projected \( i \)th vector \( \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(i) \). It should be emphasized that the choice of subspaces \( \{\mathcal{L}^{(n)}\}_n \) is independent of the index \( i \).

**Theorem 1:** (von Neumann entropy as the weak orthogonality capacity) Given a probability measure \( p \), let \( C_p(p) \) be the supremum of \( \lim \sup \inf \log L_n / n \) over all sequences \( \{\mathcal{C}^{(n)}\}_n \) that satisfy the asymptotic weak orthogonality condition. Then \( C_p(p) = H(\rho) \), where \( H(\rho) \) is the von Neumann entropy for the density operator (1).

Before proceeding to the proof, we mention a close connection between Theorem 1 and the noiseless quantum channel coding theorem [4]. Let us regard \( \mathcal{C}^{(n)} \) as a quantum random codebook. Given a vector (a quantum codeword) in \( \mathcal{C}^{(n)} \), our task is to estimate, by means of a certain measurement, which vector among \( \mathcal{C}^{(n)} \) is the actual one. Associated with a codebook \( \mathcal{C}^{(n)} \) and a subspace \( \mathcal{L}^{(n)} \) is the Gram operator

\[
\mathcal{G} := \sum_{j=1}^{L_n} |\Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(j) \rangle \langle \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(j)|.
\]

The operator \( \mathcal{G} \) is strictly positive on the subspace

\[
\hat{\mathcal{L}}^{(n)} := \text{Span}\{\Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(j); 1 \leq j \leq L_n\}.
\]

Let the operator \( \mathcal{G}^{-1} \) be the inverse of \( \mathcal{G} \) on \( \hat{\mathcal{L}}^{(n)} \) and zero on the orthogonal complement \( \hat{\mathcal{L}}^{(n)\perp} \). According to [4], we introduce a measurement \( M^{(n)} \) by

\[
M^{(n)} := \left\{ |\hat{\mu}(1)\rangle \langle \hat{\mu}(1)|, \ldots, |\hat{\mu}(L_n)\rangle \langle \hat{\mu}(L_n)|, I - \sum_{j=1}^{L_n} |\hat{\mu}(j)\rangle \langle \hat{\mu}(j)| \right\}, \quad (4)
\]

where \( \hat{\mu}(j) \) are vectors on \( \hat{\mathcal{L}}^{(n)} \) defined by

\[
\hat{\mu}(j) := \mathcal{G}^{-1/2} \Pi_{\mathcal{L}^{(n)}} \Psi^{(n)}(j) = \mathcal{G}^{-1/2} \Psi^{(n)}(j).
\]

We can regard \( M^{(n)} \) as a decoder for the codebook \( \mathcal{C}^{(n)} \), in which the \( i \)th entry \( |\hat{\mu}(i)\rangle \langle \hat{\mu}(i)|, (1 \leq i \leq L_n) \), corresponds to the \( i \)th codeword in \( \mathcal{C}^{(n)} \), and the \((L_n + 1)\)st entry to the wild card. (Note that the decoder (4) with the special choice of a subspace \( \mathcal{L}^{(n)} = \mathcal{H}^{\otimes n} \) was introduced by Holevo [5]).
The idea for adopting the decoder (4) is this: when the $i$th vector $\Psi^{(n)}(i)$ is strictly orthogonal to the other vectors in $C^{(n)}$, then the Gram operator $\mathcal{G}$ with $\mathcal{G}^{(n)} := \mathcal{H}^{\otimes n}$ is decomposed into the orthogonal direct sum

$$\mathcal{G} = |\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)| \oplus \sum_{j \neq i} |\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)|,$$

so that

$$\mathcal{G}^{-1/2} = |\Psi^{(n)}(i)\rangle\langle\Psi^{(n)}(i)| \oplus \left( \sum_{j \neq i} |\Psi^{(n)}(j)\rangle\langle\Psi^{(n)}(j)| \right)^{-1/2},$$

and

$$\hat{\mu}(i) = \mathcal{G}^{-1/2}\Psi^{(n)}(i) = \Psi^{(n)}(i).$$

As a consequence, the decoding error probability $P_{e}^{(n)}(i)$ for the $i$th codeword $\Psi^{(n)}(i)$ by the decoder (4) is

$$P_{e}^{(n)}(i) = 1 - \text{Tr}[\Psi^{(n)}(i)\langle\Psi^{(n)}(i)|\hat{\mu}(i)\rangle\langle\hat{\mu}(i)|] = 0.$$

Thus it is expected that, when the $i$th vector is almost orthogonal to the other vectors in $C^{(n)}$, the decoding error probability $P_{e}^{(n)}(i)$ will be small. In fact, this expectation is verified by the following

**Lemma 2:** If $|\hat{g}_{ii}^{(n)}|^2 > 1 - \varepsilon$ and $\hat{y}_{i}^{(n)} < \varepsilon$ hold for some $i$ and $0 < \varepsilon < 1$, then the decoding error probability $P_{e}^{(n)}(i)$ for the $i$th codeword $\Psi^{(n)}(i)$ by the decoder (4) is upper bounded by $\frac{3}{2}\varepsilon$.

**Proof:** See Appendix B. \hfill \blacksquare

**Corollary 3:** If $\{C^{(n)}\}_n$ satisfies asymptotic weak orthogonality condition, then $E[P_{e}^{(n)}] \to 0$ as $n \to \infty$, where $P_{e}^{(n)}$ is the average decoding error probability for the code $(C^{(n)}, M^{(n)})$, and $E[\cdot]$ denotes the expectation.

**Proof:** Let the event $E_{i}^{(n)}(\varepsilon)$ be defined by

$$E_{i}^{(n)}(\varepsilon) := \{|\hat{g}_{ii}^{(n)}|^2 > 1 - \varepsilon \text{ and } \hat{y}_{i}^{(n)} < \varepsilon\}.$$

By the assumption of asymptotic weak orthogonality, for all $\varepsilon > 0$ and $\delta > 0$, there is an $N$ such that for all $n \geq N$ and $i$, $P(E_{i}^{(n)}(\varepsilon)) > 1 - \delta$ holds. Then

$$E[P_{e}^{(n)}(i)] = E[P_{e}^{(n)}(i); E_{i}^{(n)}(\varepsilon)] + E[P_{e}^{(n)}(i); E_{i}^{(n)}(\varepsilon)^c]$$

$$\leq E[P_{e}^{(n)}(i); E_{i}^{(n)}(\varepsilon)] + P(E_{i}^{(n)}(\varepsilon)^c) < \frac{3}{2}\varepsilon + \delta.$$
Here, $E[X; A] := \int_A X dP$, and Lemma 2 is used in the last inequality. Since this upper bound is independent of $i$, we have

$$E[P_e^{(n)}] = \frac{1}{L_n} \sum_{i=1}^{L_n} E[P_e^{(n)}(i)] < \frac{3}{2} \varepsilon + \delta.$$ 

This completes the proof.

Theorem 1 and Corollary 3 clarify why the decoder of the type (4) has fitted to the random coding technique in the proof of the direct part of the noiseless quantum channel coding theorem [4]. The notion of asymptotic weak orthogonality for the random codebook $C^{(n)}$ thus explicates the physical implication of the probabilistic distinguishability among codewords in the quantum channel coding problem as well as the geometrical mechanism behind the decoder (4).

**Proof of Theorem 1:** We first prove the direct part $C_w(p) \geq H(\rho)$. Fix an arbitrarily small positive constant $\delta$ and, for each $n$, let $L_n$ be such that $L_n < e^{n(H(\rho) - 4\delta)}$. We show that there is a sequence $\{L^{(n)}\}_n$ of subspaces for which the asymptotic weak orthogonality conditions (i) and (ii) hold for all $i$. The idea of the proof is similar to [4]: we take $L^{(n)}$ to be the $\delta$-typical subspace $\Lambda_\delta^{(n)}$ with respect to the density $\rho$. (For the reader’s convenience, the definition and the basic properties of the $\delta$-typical subspace are summarized in Appendix C.) Since

$$\hat{g}_{ii}^{(n)} = \langle \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(i) | \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(i) \rangle = \text{Tr} \rho^{\otimes n} \Pi_{\Lambda_\delta^{(n)}},$$

we have

$$E[\hat{g}_{ii}^{(n)}] = \text{Tr} \rho^{\otimes n} \Pi_{\Lambda_\delta^{(n)}} > 1 - \delta$$

for all $i$ and all sufficiently large $n$ (see (9)). This proves (i). On the other hand, for all $j \neq i$,

$$|\hat{g}_{ij}^{(n)}|^2 = |\langle \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(i) | \Pi_{\Lambda_\delta^{(n)}} \Psi^{(n)}(j) \rangle|^2$$

$$= \text{Tr} \Pi_{\Lambda_\delta^{(n)}} |\Psi^{(n)}(i)\rangle \langle \Psi^{(n)}(i)| \Pi_{\Lambda_\delta^{(n)}} |\Psi^{(n)}(j)\rangle \langle \Psi^{(n)}(j)| \Pi_{\Lambda_\delta^{(n)}},$$

so that

$$E[\hat{g}_{ij}^{(n)}]^2 = \text{Tr} \Pi_{\Lambda_\delta^{(n)}} \rho^{\otimes n} \Pi_{\Lambda_\delta^{(n)}} \rho^{\otimes n} \Pi_{\Lambda_\delta^{(n)}} = \text{Tr} (\rho^{\otimes n})^2 \Pi_{\Lambda_\delta^{(n)}} \leq e^{-n(H(\rho) - 3\delta)},$$

(see (10)), and

$$E[\hat{Y}_{\delta}^{(n)}] \leq (L_n - 1) e^{-n(H(\rho) - 3\delta)} < e^{-n\delta}.$$
Thus $\hat{Y}_i^{(n)} \to 0$ in $L^1$ for all $i$, proving (ii). This completes the proof of the direct part $C_w(p) \geq H(\rho)$.

We next prove the converse part $C_w(p) \leq H(\rho)$. Let $X$ be the random variable uniformly distributed over $C^{(n)}$ and let $Y$ be the random variable representing the outcome of the corresponding decoder $M^{(n)}$ defined by (4). Then by virtue of Fano’s inequality,

$$
\log 2 + P_c^{(n)} \log L_n \geq H(X | Y) = H(X) - I(X : Y)
$$

$$
= \log L_n - \frac{1}{L_n} \sum_{j=1}^{L_n} D_{M^{(n)}} \left( |\Psi^{(n)}(j)\rangle \langle \Psi^{(n)}(j)| \right) \left( \left\| \frac{1}{L_n} \sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle \langle \Psi^{(n)}(k)| \right\| \right)
$$

$$
\geq \log L_n - \frac{1}{L_n} \sum_{j=1}^{L_n} D \left( |\Psi^{(n)}(j)\rangle \langle \Psi^{(n)}(j)| \right) \left( \left\| \frac{1}{L_n} \sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle \langle \Psi^{(n)}(k)| \right\| \right)
$$

$$
= \log L_n - H \left( \frac{1}{L_n} \sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle \langle \Psi^{(n)}(k)| \right).
$$

Here $D(\sigma \| \tau) := \text{Tr} \sigma (\log \sigma - \log \tau)$ is the quantum relative entropy between the quantum states $\sigma$ and $\tau$ (with $\text{supp} \sigma \subset \text{supp} \tau$), and $D_M(\sigma \| \tau)$ denotes the classical Kullback-Leibler divergence between the probability distributions $p(\cdot) := \text{Tr} \sigma M(\cdot)$ and $q(\cdot) := \text{Tr} \tau M(\cdot)$ over the outcomes of the measurement $M$. (See [3] for notations.) The second inequality is due to the familiar monotonicity relation of the relative entropy (Theorem 1.5 [7]). Now taking the expectation for the above inequality, and using the concavity of the von Neumann entropy, we have

$$
\log 2 + E[P_c^{(n)}] \log L_n \geq \log L_n - E \left[ H \left( \frac{1}{L_n} \sum_{k=1}^{L_n} |\Psi^{(n)}(k)\rangle \langle \Psi^{(n)}(k)| \right) \right]
$$

$$
\geq \log L_n - H \left( \frac{1}{L_n} \sum_{k=1}^{L_n} E \left[ |\Psi^{(n)}(k)\rangle \langle \Psi^{(n)}(k)| \right] \right)
$$

$$
= \log L_n - H(\rho^{\otimes n}) = \log L_n - nH(\rho).
$$

Therefore

$$
(1 - E[P_c^{(n)}]) \frac{\log L_n}{n} \leq H(\rho) + \frac{\log 2}{n}.
$$

Thus in order to assure the asymptotic weak orthogonality (so that $E[P_c^{(n)}] \to 0$ as $n \to \infty$ by Corollary 3), $\lim sup_n \log L_n / n$ must be less than or equal to $H(\rho)$. This completes the proof of the converse part $C_w(p) \leq H(\rho)$.
3. Strong Orthogonality

If the vectors in $\mathcal{C}^{(n)}$ are mutually strictly orthogonal, then the Gram matrix $G^{(n)} = [g_{ij}^{(n)}]$, where $g_{ij}^{(n)}$ is the inner product (2), is reduced to the identity. Therefore if they are mutually “almost” orthogonal, the Gram matrix is expected to be close to the identity. This observation prompts us to define the strong orthogonality as follows. Given $\mathcal{C}^{(n)}$, let the random variable $Z^{(n)}$ be defined by the squared sum of off-diagonal elements of $G^{(n)}$, i.e.,

$$Z^{(n)} := \sum_{i=1}^{L_n} \sum_{j \neq i} |g_{ij}^{(n)}|^2.$$  

We say that a sequence $\{C^{(n)}\}$ satisfies the asymptotic strong orthogonality condition if the following two conditions are satisfied:

(i) $Z^{(n)} \to 0$ in probability,

(ii) the sequence $\{Z^{(n)}\}$ is uniformly integrable.

**Theorem 4:** (Quantum Rényi entropy as the strong orthogonality capacity) Given a probability measure $p$, let $C_s(p)$ be the supremum of $\limsup_{n \to \infty} \log L_n/n$ over all sequences $\{C^{(n)}\}$ that satisfy asymptotic strong orthogonality condition. Then $C_s(p) = \frac{1}{2} H_2(\rho)$, where $H_2(\rho)$ is the quantum Rényi entropy of degree 2 for the density operator (1).

**Proof:** For $i \neq j$,

$$|g_{ij}^{(n)}|^2 = |\langle \Psi^{(n)}(i) | \Psi^{(n)}(j) \rangle|^2 = \text{Tr} (|\Psi^{(n)}(i)\rangle\langle \Psi^{(n)}(i)|)(|\Psi^{(n)}(j)\rangle\langle \Psi^{(n)}(j)|),$$

so that

$$E|g_{ij}^{(n)}|^2 = \text{Tr} (\rho^\otimes n)^2 = (\text{Tr} \rho^2)^n = e^{-n H_2(\rho)},$$

and

$$E[Z^{(n)}] = L_n(L_n - 1) e^{-n H_2(\rho)}.$$  

Let $\delta > 0$ be an arbitrarily small positive constant. If $L_n < e^{n(H_2(\rho)/2 - \delta)}$ then $E[Z^{(n)}] \to 0$, and if $L_n > e^{n(H_2(\rho)/2 + \delta)}$ then $E[Z^{(n)}] \to \infty$. This completes the proof.

Several remarks are in order. If $\{C^{(n)}\}$ satisfies condition (i) for asymptotic strong orthogonality, then in a similar way to Lemma 2, it can be shown that the decoding error probabilities $\{P_e^{(n)}(i)\}_{1 \leq i \leq L_n}$ by the decoder (4) with $\mathcal{L}^{(n)} := \mathcal{H}^\otimes n$ exhibit $\sum_{i=1}^{L_n} P_e^{(n)}(i) \to 0$ in $L^1$ as $n \to \infty$. (Compare this with Corollary 3.) The notion of asymptotic strong orthogonality thus leads to a new, strong type of probabilistic distinguishability.
in quantum measurement theory. It is not clear whether there is a coding theorem in which this strong distinguishability plays a pivotal role. A related question whether the converse part of Theorem 4 holds without the uniform integrability condition (ii) is also still open.

4. Conclusions
We have shown that asymptotic orthogonalities of random vectors lead us to new, geometrical, characterizations of the von Neumann entropy and the quantum Rényi entropy of degree 2. These characterizations are closely related to the distinguishability of the vectors by quantum mechanical measurements. In particular, a mechanism behind the random coding technique for the noiseless quantum channel coding theorem was clarified.

Appendix A. Remark on Condition (3)
In this appendix, we show that condition (3) does not characterize the von Neumann entropy. Let \( \phi_0, \phi_1 \) be unit vectors in \( \mathcal{H} \) with \( a := |\langle \phi_0 | \phi_1 \rangle|^2 < 1 \), and let \( p \) be the probability measure on \( \mathcal{H} \) such that \( p(\phi_0) = p(\phi_1) = \frac{1}{2} \). The nonzero eigenvalues of the corresponding density operator \( \rho = \frac{1}{2} |\phi_0 \rangle \langle \phi_0 | + \frac{1}{2} |\phi_1 \rangle \langle \phi_1 | \) are \( (1 \pm \sqrt{a})/2 \), so the quantum Rényi entropy of degree 2 is

\[
h_2 := H_2(\rho) = -\log \frac{1+a}{2}.
\]

Let \( \{X_k(i)\}_{ki} \) be \( \{\phi_0, \phi_1\} \)-valued random variables i.i.d. with respect to \( p \), and let \( \Psi^{(n)}(i) = X_1(i) \otimes \cdots \otimes X_n(i) \). The squared norm of the inner product \( g_{ij}^{(n)} \) then becomes

\[
|g_{ij}^{(n)}|^2 = \prod_{k=1}^{n} |\langle X_k(i) | X_k(j) \rangle|^2 = a^{N_{ij}},
\]

where \( N_{ij} \) is the number of indices \( k \) for which \( X_k(i) \neq X_k(j) \).

Now fix a number \( i \) arbitrarily. Then it is easily shown that for each \( n \), \( \{|g_{ij}^{(n)}|^2; j \in \mathcal{N}, j \neq i\} \) are i.i.d. random variables, each taking the value \( a^\ell \) with probability \( \binom{n}{\ell} 2^{-n} \), where \( \ell = 0, \ldots, n \). In particular, they have the expectation

\[
m^{(n)} := E \left[ |g_{ij}^{(n)}|^2 \right] = \left( \frac{1+a}{2} \right)^n = e^{-nh_2},
\]

and the variance

\[
v^{(n)} := V \left[ |g_{ij}^{(n)}|^2 \right] = \left( \frac{1+a^2}{2} \right)^n - \left( \frac{1+a}{2} \right)^{2n}.
\]
We claim the following.

**Proposition 5:** Let \( \varepsilon \) be a positive constant and let

\[
Y_i^{(n)} = \sum_{j \neq i}^{L_n} |g_{ij}^{(n)}|^2.
\]

If \( L_n < e^{n(h_2 - \varepsilon)} \), then \( Y_i^{(n)} \) converges to 0 in probability as \( n \to \infty \), and if \( L_n > e^{n(h_2 + \varepsilon)} \), then \( Y_i^{(n)} \) does not.

**Proof:** Assume first that \( L_n < e^{n(h_2 - \varepsilon)} \). Then

\[
E[Y_i^{(n)}] = (L_n - 1) m^{(n)} < e^{-n \varepsilon} \to 0 \text{ as } n \to \infty,
\]

proving that \( Y_i^{(n)} \to 0 \) in probability. To prove the second part, we use the following inequality:

\[
P\left( Y_i^{(n)} < \frac{1}{2}(L_n - 1) m^{(n)} \right) < \frac{4}{(L_n - 1) m^{(n)}}, \tag{5}
\]

which is verified as follows.

\[
P\left( Y_i^{(n)} < \frac{(L_n - 1)m^{(n)}}{2} \right) \leq P\left( \left| Y_i^{(n)} - (L_n - 1)m^{(n)} \right| > \frac{(L_n - 1)m^{(n)}}{2} \right)
\]

\[
< \left( \frac{2}{(L_n - 1)m^{(n)}} \right)^2 V[Y_i^{(n)}] = \frac{4}{(L_n - 1)m^{(n)}} \cdot \frac{v^{(n)}}{m^{(n)}} < \frac{4}{(L_n - 1)m^{(n)}}.
\]

Now assume that \( L_n > e^{n(h_2 + \varepsilon)} \). Then \( (L_n - 1)m^{(n)} > e^{n \varepsilon} - m^{(n)} \to \infty \) as \( n \to \infty \). This fact and the inequality (5) together prove that \( Y_i^{(n)} \) does not converge to 0 in probability.

According to Proposition 5, it is \( h_2 \) (the quantum Rényi entropy of degree 2) that characterizes the asymptotic behavior of \( Y_i^{(n)} \) in this example. As a consequence, one cannot characterize the von Neumann entropy as the capacity for condition (3) in general.

**Remark 6:** If the definition of the asymptotic weak orthogonality is such that \( Y_i^{(n)} \to 0 \) in probability and the sequence \( \{Y_i^{(n)}\}_n \) is uniformly integrable for all \( i \), then it can be shown in a similar way to Theorem 4 that \( C_w(p) = H_2(\rho) \) for a general probability measure \( p \).

**Appendix B. Proof of Lemma 2**

Due to the symmetry, it suffices to consider the case when \( i = 1 \). Let \( \{\hat{e}_k\}_k \) be a complete orthonormal system (CONS) of the finite dimensional subspace \( \hat{L}^{(n)} \) with

\[
\hat{e}_1 := \frac{\Pi_{\hat{L}^{(n)}} \Psi^{(n)}(1)}{\|\Pi_{\hat{L}^{(n)}} \Psi^{(n)}(1)\|}.
\]
The (1,1)th matrix element of $G^{-1/2}$ with respect to the CONS $\{\hat{e}_k\}_k$ is

$$
(G^{-1/2})_{11} := \langle \hat{e}_1 | G^{-1/2} \hat{e}_1 \rangle = \frac{\langle \Pi_{L(n)} \Psi^{(n)}(1) | G^{-1/2} \Pi_{L(n)} \Psi^{(n)}(1) \rangle}{\| \Pi_{L(n)} \Psi^{(n)}(1) \|^2} \\
= \frac{\langle \Psi^{(n)}(1) | G^{-1/2} \Psi^{(n)}(1) \rangle}{\| \Pi_{L(n)} \Psi^{(n)}(1) \|^2}.
$$

(6)

The error probability $P_e^{(n)}(1)$ for the first codeword $\Psi^{(n)}(1)$ in $\mathcal{C}^{(n)}$ with respect to the decoder (4) is evaluated as

$$
P_e^{(n)}(1) = 1 - |\langle \Psi^{(n)}(1) | \hat{\mu}(1) \rangle|^2 = 1 - |\langle \Psi^{(n)}(1) | G^{-1/2} \Psi^{(n)}(1) \rangle|^2 \\
= 1 - \| \Pi_{L(n)} \Psi^{(n)}(1) \|^4 |(G^{-1/2})_{11}|^2 \leq 1 - \| \Pi_{L(n)} \Psi^{(n)}(1) \|^4 (G_{11})^{-1}. (7)
$$

Here $G_{11}$ stands for the (1,1)th matrix element of $G$, and we have used (6) and the inequality

$$
(G^{-1/2})_{11} \geq (G_{11})^{-1/2},
$$

which is verified by Lemma 7 below. On the other hand,

$$
G_{11} = |\langle \hat{e}_1 | \Pi_{L(n)} \Psi^{(n)}(1) \rangle|^2 + \sum_{j \geq 2} |\langle \hat{e}_1 | \Pi_{L(n)} \Psi^{(n)}(j) \rangle|^2 \\
= \| \Pi_{L(n)} \Psi^{(n)}(1) \|^2 + \frac{1}{\| \Pi_{L(n)} \Psi^{(n)}(1) \|^2} \sum_{j \geq 2} |\langle \Pi_{L(n)} \Psi^{(n)}(1) | \Pi_{L(n)} \Psi^{(n)}(j) \rangle|^2 \\
= \| \Pi_{L(n)} \Psi^{(n)}(1) \|^2 \left( 1 + \frac{\sum_{j \geq 2} |\langle \Pi_{L(n)} \Psi^{(n)}(1) | \Pi_{L(n)} \Psi^{(n)}(j) \rangle|^2}{\| \Pi_{L(n)} \Psi^{(n)}(1) \|^4} \right) \\
< \| \Pi_{L(n)} \Psi^{(n)}(1) \|^2 \left( 1 + \frac{\varepsilon}{1 - \varepsilon} \right) = \| \Pi_{L(n)} \Psi^{(n)}(1) \|^2 \frac{1}{1 - \varepsilon}. (8)
$$

Substituting (8) into (7), we have

$$
P_e^{(n)}(1) < 1 - \| \Pi_{L(n)} \Psi^{(n)}(1) \|^2 (1 - \varepsilon) < 1 - (1 - \varepsilon)^{3/2} < \frac{3}{2} \varepsilon.
$$

This completes the proof of Lemma 2.

**Lemma 7:** Let $A = [A_{ij}]$ be a strictly positive Hermitian matrix. Then*

$$
(A^{-1/2})_{11} \geq (A_{11})^{-1/2}.
$$

*More generally, we can prove that $(A^m)_{11} \geq (A_{11})^m$ for $m < 0$ or $m > 1$, and $(A^m)_{11} \leq (A_{11})^m$ otherwise.
Proof: Let 

\[ A = \sum_k \lambda_k E_k \]

be the spectral decomposition. Then 

\[ A^{-1/2} = \sum_k (\lambda_k)^{-1/2} E_k, \]

so that 

\[
(A^{-1/2})_{11} = \sum_k (\lambda_k)^{-1/2} \langle e_1 | E_k e_1 \rangle \geq \left( \sum_k \lambda_k \langle e_1 | E_k e_1 \rangle \right)^{-1/2} = (A_{11})^{-1/2}.
\]

Here we have used Jensen’s inequality and the fact that \( \langle e_1 | E_k e_1 \rangle \geq 0 \) for all \( k \) and \( \sum_k \langle e_1 | E_k e_1 \rangle = 1 \).

Appendix C. Typical Subspaces

In this appendix, we give a brief account of the so-called typical subspace (cf. [9, 6, 4]). Given a density operator \( \rho \) on \( \mathcal{H} \), let 

\[ \rho = \sum_{j \in J} \lambda_j E_j \]

be a Schatten decomposition, where \( \lambda_j > 0 \) for all \( j \in J \) and \( \sum_j \lambda_j = 1 \). Note that \( \lambda := (\lambda_1, \lambda_2, \ldots) \) is naturally regarded as a probability distribution on the index set \( J \) such that \( \lambda(j) = \lambda_j \). A Schatten decomposition of \( \rho^\otimes n \) is given by

\[ \rho^\otimes n = \sum_{(j_1, \ldots, j_n) \in J^n} (\lambda_{j_1} \cdots \lambda_{j_n})(E_{j_1} \otimes \cdots \otimes E_{j_n}). \]

Obviously, the eigenvalues of \( \rho^\otimes n \) form a probability distribution \( \lambda^n \), the i.i.d. extension of \( \lambda \), on the set \( J^n \).

Given a density operator \( \rho \) and a positive constant \( \delta \), an eigenvalue \( (\lambda_{j_1} \cdots \lambda_{j_n}) \) of \( \rho^\otimes n \) is called \( \delta \)-typical if the sequence \( j_1, \ldots, j_n \) of indices is \( \delta \)-typical (p. 51 [2]) with respect to the probability distribution \( \lambda^n \), that is, if

\[ e^{-n(H(\rho)+\delta)} \leq \lambda^n(j_1, \ldots, j_n) \leq e^{-n(H(\rho)-\delta)}. \]

It follows that for all sufficiently large \( n \),

(a) \( e^{-n(H(\rho)+\delta)} \leq (\text{a } \delta \text{-typical eigenvalue}) \leq e^{-n(H(\rho)-\delta)} \),

(b) (the sum of \( \delta \)-typical eigenvalues) \( > 1 - \delta \),

(c) \( (1 - \delta)e^{n(H(\rho)-\delta)} \leq (\text{the number of } \delta \text{-typical eigenvalues}) \leq e^{n(H(\rho)+\delta)}. \)
Here (a) is a direct consequence of the definition, and (b) and (c) follow from the asymptotic equipartition property (Theorem 3.1.2 [2]).

Let \( \Lambda_\delta^{(n)}(\subset \mathcal{H} \otimes \mathcal{H}^n) \) be the linear span of such eigenvectors of \( \rho \otimes \rho^n \) that correspond to \( \delta \)-typical eigenvalues. The subspace \( \Lambda_\delta^{(n)} \) is called \( \delta \)-typical with respect to the density \( \rho \). Let \( \Pi_{\Lambda_\delta^{(n)}} \) be the projection operator onto the \( \delta \)-typical subspace \( \Lambda_\delta^{(n)} \). Clearly the operators \( \rho \) and \( \Pi_{\Lambda_\delta^{(n)}} \) commute. And it follows immediately from (a)–(c) that, for all sufficiently large \( n \),

\[
\Tr \rho \otimes \rho^n \Pi_{\Lambda_\delta^{(n)}} > 1 - \delta, \tag{9}
\]

\[
\Tr (\rho \otimes \rho^n)^2 \Pi_{\Lambda_\delta^{(n)}} \leq \left( e^{-n(H(\rho)-\delta)} \right)^2 e^{n(H(\rho)+\delta)} = e^{-n(H(\rho)-3\delta)}. \tag{10}
\]

**Acknowledgments**

The author wishes to thank Michael Camarri and Hidetoshi Shimokawa for their helpful comments. He also acknowledges support of the Telecommunications Advancement Foundation.

**References**


PART II
Quantum Cramér-Rao Bound in Mixed States Model

Chap. 8: H. Nagaoka
“A new approach to Cramér-Rao bounds for quantum state estimation” ....................... 100

Chap. 9: H. Nagaoka
“On Fisher information of quantum statistical models” ....................... 113

Chap. 10: H. Nagaoka
“On the parameter estimation problem for quantum statistical models” ....................... 125

Chap. 11: H. Nagaoka
“A generalization of the simultaneous diagonalization of Hermitian matrices and its relation to quantum estimation theory” ....................... 133

Chap. 12: M. Hayashi
“A linear programming approach to attainable Cramér-Rao type bounds” ....................... 150

Chap. 13: M. Hayashi and K. Matsumoto
“Statistical model with measurement degree of freedom and quantum physics” ....................... 162

Chap. 14: M. Hayashi
“Asymptotic quantum theory for the thermal states family” ....................... 170

Chap. 15: R. D. Gill and S. Massar
“State estimation for large ensembles” ....................... 178
CHAPTER 7

Introduction to Part II

1. Main Issue in Quantum Estimation

The aim of quantum statistical estimation theory is to estimate the true density matrix of the quantum system of interest through suitable quantum measurements. Its process is divided into two parts: One is the measurement process, in which we perform a quantum measurement on the quantum system of interest. The other is data manipulation, in which we estimate the true parameter from the obtained outcomes. Thus, for our purpose, we choose these two processes to be optimal. While the latter process can be treated as a problem in the classical estimation theory, the selection of the former is treated specially in the quantum case. Therefore, the choice of the measurement is the main issue in quantum estimation.

2. Asymptotic Theory of Classical Estimation

Here, let us briefly review estimation theory in classical statistics. Statistical researchers usually assume that the true probability distribution belongs to a certain parametric subset of probability distributions, which we refer to as a family or a family of probability distributions, that is described by \( \{ p_\theta | \theta \in \Theta \} \), where \( \Theta \) is the parameter space. The family of probability distributions is chosen on their prior knowledge. This assumption is effective for reducing the estimation error, and is a standard method in classical statistics.

Under the one-parameter case, the estimation error is usually measured by the mean square error (MSE), which is defined as the expectation of the square of the difference between the true parameter and the estimated parameter under the true probability distribution. When the amount of obtained data is large enough, the MSE of a suitable estimator is approximately proportional to the inverse of the number of observations. In such an asymptotic case, it is natural to focus on its coefficient; and the optimal coefficient equals the Fisher information, which is defined as the expectation of the squared logarithmic derivative of the probability density with respect to the parameter.
In the multi-parameter case, some researchers focus on the sum of the MSEs of the respective parameters. In the asymptotic case, that of the optimal estimator is also proportional to the inverse of the number of observations, and its coefficient equals the trace of the inverse of the *Fisher information matrix*, which can be regarded as a multi-parameter extension of Fisher information.

3. Attainable Quantum Cramér-Rao Type Bound

In the quantum case, similarly to the classical case, it is often known that the true quantum state belongs to a certain parametric subset of densities, referred to as a *quantum state family*, which is described by \( \{ \rho_\theta | \theta \in \Theta \} \). Since the above-mentioned general framework has been established in the asymptotic theory of classical estimation, it is very natural to extend it to the quantum setting.

In this setting, when we fix a quantum measurement \( M \), the outcome obeys a probability distribution \( P^M_\rho \). Thus, the choice of the quantum measurement \( M \) implies the choice of the family of probability distributions \( \{ P^M_\rho | \theta \in \Theta \} \). Hence, when we perform the same measurement \( M \) on the separate systems in the quantum i.i.d. case, the MSE of the best classical data manipulation is asymptotically characterized by Fisher information (or Fisher information matrix) of the probability distribution family \( \{ P^M_\rho | \theta \in \Theta \} \). Therefore, the central issue of quantum estimation is the minimization problem of the inverse of Fisher information (or the trace of the inverse of Fisher information matrix) over the family of probability distributions defined by the quantum measurement. Its minimum value is called the *most informative Cramér-Rao bound*, the *attainable Cramér-Rao type bound*, or the *non-quantum-correlational Cramér-Rao type bound*, which is simplified to the CR bound in the following. Indeed, it is possible to replace the minimization problem by the minimization of the trace of the covariance matrix of a locally unbiased estimator. This is because the former minimum is equivalent to the latter minimum, as was shown by Nagaoka [Chap. 8]*.

As the first analysis of the CR bound, Helstrom [II-1] introduced *symmetric logarithmic derivative* (SLD) and *SLD Fisher information* as quantum analogues of logarithmic derivative and Fisher information, respectively. Using these, he solved this optimization problem in the one-

*While the paper [Chap. 9] is an older paper than [Chap. 8], the paper [Chap. 8] is treated as the first paper in Part II because Chap. 8 is more introductory.*
parameter case. However, the optimization remained open in the multi-parameter case. The selected papers [Chaps. 5, 8, 9, 12, 15, 16] treated this problem, and calculated the CR bound in some specific cases. Here, we review the history of the CR bound in the multi-parameter case. Extensive progress in this case was first made by Yuen and Lax [II-2]. They solved this problem in the estimation problem of the coherent states family in thermal noise. Following this result, Holevo [0-2] expanded their result to the estimation problem of the expectation parameter of quantum Gaussian state family. His method is based on right logarithmic derivative (RLD) and another quantum analogue of Fisher information matrix which is called the RLD Fisher information matrix. He also obtained a lower bound to the CR bound, which can be applied in much great generality, and is called the Holevo bound.

However, the researches until 1980 did not focus on the relation between asymptotic theory and the latter minimum value. They were only concerned with the minimization under the unbiasedness condition or the locally unbiasedness condition. Since Nagaoka [Chap. 8] pointed out this relation in the middle of the 1980s, its asymptotic aspect has been studied. Nagaoka [Chap. 8] derived a more simple expression for the Holevo bound and a better lower bound in the two-parameter case, which is called the Nagaoka bound. Nagaoka [Chap. 11] calculated the CR bound in any two-parameter family in the quantum two-level system by showing that the Nagaoka bound is achievable in these families. He treated the simultaneous measurement of two observables [Chap. 11], which was generalized to that of multi-observables by Hayashi. Hayashi [Chap. 12] calculated the CR bound of any family in the quantum two-level system [Chap. 12]. His method is based on the duality theorem of linear programming, and can be applied to the three-parameter case as well as the two-parameter case. After this progress, Gill and Massar [Chap. 15] independently obtained the CR bound of any family in the quantum two level system. (More details on the mathematical statistical side of their approach is discussed in another paper [II-3].) Their approach is completely different from the preceding ones, and is simpler than the previous approaches.

†Only Young’s paper [II-7] treated this relation.
4. Attainable Quantum Cramér-Rao Type Bound with Quantum Correlation

In the previous section, we discussed the case where we do not use quantum correlation on the tensor product space in the measuring apparatus. In the following, we focus on the possible advantage of collective measurements§, i.e., we consider whether using quantum correlation in the measuring process can improve the CR bound or not. It is easily verified that the collective measurement cannot improve this bound in the one-parameter case because the SLD Fisher information on the $n$-tensor product space equals $n$ times the SLD Fisher information. But, the situation is completely changed in the multi-parameter case as follows: Hayashi discussed this problem in the quantum two-level system [II-5] and the simultaneous estimation problem of the displacement parameter and thermal noise parameter for a family of coherent states in thermal noise [Chap. 14]. Regarding the quantum two-level system, the tensor product states can be approximated by the coherent states family in the thermal noise. Thus, the results of the latter can be applied to the former case. Moreover, using group the representation theory, Matsumoto treated the CR bound in any family in the finite-dimensional system¶. Using the quantum central limit theorem, Hayashi obtained the same results in a more general setting [II-9]. In addition, any adaptive POVM cannot improve the Fisher information matrix based on a single POVM. This quantum two-level system case was proven by Gill and Massar [Chap. 15], and its general case by Hayashi [II-6]. Thus, for further improvement, we require a POVM that is not contained by separable POVMs.

Furthermore, Tsuda and Matsumoto [I-6] extended some of the above results of Quantum Cramér-Rao bounds to the non-differentiable settings.

5. Global Attainability of CR Bound

It is a serious problem, however, that the optimal measurement realizing the CR bound depends on the true state, in general. Nagaoka focused on a family wherein such a measurement is independent of the true state, and called such a family quasi-classical. He obtained several equivalent conditions for

§ We use the term ‘collective measurement’ in the sense of quantum measurement in whose process we use quantum correlation on the tensor product space, while some experimental researchers use it in the sense of the measurement concerning the ensemble average.

quasi-classicality [Chap. 9], whereas Young obtained similar results [II-7]. However, even if the state family is not quasi-classical, by adaptively choosing the measurement, we can asymptotically achieve the CR bound. This possibility was pointed by Nagaoka [Chap. 10], and independently proven by Hayashi & Matsumoto [Chap. 13] and Gill & Massar [Chap. 15] \[\text{\cite{II-7}}\]. This result implies that in the one-parameter case, the optimal error with quantum correlation can be asymptotically attained by adaptively choosing the measurement on the single system. Indeed, Fischer and Freyberger [II-10] proposed an experimental scheme of adaptive measurement, and Hannemann \textit{et al.} [II-11] experimentally realized their proposal. Moreover, Nagaoka also focused on a quantum analogue of an efficient estimator [Chap. 9], and discussed a quantum extension of information geometry [Chap. 10] (This discussion is also contained in Amari and Nagaoka [II-12]).

Finally, we should remark that there are other results regarding the CR bound in the pure states case. These will be discussed in the next part.

6. Mathematical Formulation in Classical Estimation

6.1. \textit{Cramér-Rao Inequality}

For the reader’s convenience, we will review the asymptotic theory in classical estimation, more formally.

Assume that \( n \) data \( \bar{\omega}_n \equiv (\omega_1, \ldots, \omega_n) \) are independently generated subject to an unknown probability distribution, which belongs to a one-parametric probability family \( \{p_\theta | \theta \in \mathbb{R}\} \). Then, the estimator is given by a function \( \hat{\theta}_n \) from the \( n \) observations to the real number. Usually, the estimation error is evaluated by the MSE, which is defined by

\[
\text{MSE}_\theta(\hat{\theta}_n) \equiv \sum_{\bar{\omega}_n}(\theta - \hat{\theta}_n(\bar{\omega}_n))^2 p^n_\theta(\bar{\omega}_n), \tag{1}
\]

where \( p^n_\theta \) is the \( n \)-fold product density obtained from \( p_\theta \). When the estimator \( \hat{\theta}_n \) satisfies the following condition, it is called \textit{unbiased}:

\[
\theta = \sum_{\bar{\omega}_n} \hat{\theta}_n(\bar{\omega}_n)p^n_\theta(\bar{\omega}_n), \quad \forall \theta \in \mathbb{R}. \tag{2}
\]

\[\text{\cite{II-7}}\] The Gill-Massar paper was inspired by Barndorff-Nielsen and Gill [II-8], which roughly sketched how asymptotically optimal adaptive estimation in the one-parameter case could be done, in general.
Any unbiased estimator $\hat{\theta}_n$ satisfies the inequality:

$$\text{MSE}_\theta(\hat{\theta}_n) \geq \frac{1}{nJ_\theta}$$

which is called the Cramér-Rao inequality, and $J_\theta$ is the Fisher information defined as

$$J_\theta \equiv \sum_\omega \left( \frac{d \log p_\theta(\omega)}{d \theta} \right)^2 p_\theta(\omega).$$

If an unbiased estimator attains the above equality at all points, it is called efficient. Indeed, it is easily verified that the following three conditions are equivalent for a family $\{p_\theta|\theta \in \mathbb{R}\}$:

(i) There exists an efficient estimator $X$ for the family.

(ii) There exists an efficient estimator $X^n$ for the $n$-fold family $\{p_\theta^n|\theta \in \mathbb{R}\}$.

(iii) The family $\{p_\theta|\theta \in \mathbb{R}\}$ has another parameter $\tilde{\theta}$ such that

$$p_{\tilde{\theta}}(\omega) = \exp(\tilde{\theta}X(\omega) - \psi(\tilde{\theta})),$$

$$\psi(\tilde{\theta}) \equiv \log \sum_\omega \exp(\tilde{\theta}X(\omega)).$$

Especially, if the above condition is satisfied, the family $\{p_\theta|\theta \in \mathbb{R}\}$ is called exponential. Moreover, when the condition (iii) holds, we have the following relations

$$\theta = \frac{d\psi(\tilde{\theta})}{d\theta}, \quad \tilde{\theta} = \frac{d\phi(\theta)}{d\theta},$$

where $\phi(\theta) \equiv \max_{\tilde{\theta}} \tilde{\theta}\theta - \psi(\tilde{\theta})$.

6.2. Asymptotic Setting

When the number $n$ of obtained data is large, the maximum likelihood estimator $\hat{\theta}_n^{\text{ML}}(\vec{\omega}_n)$ is almost optimal in the following sense: The maximum likelihood estimator satisfies

$$\lim_{n \to \infty} n\text{MSE}_\theta(\hat{\theta}_n^{\text{ML}}) = \frac{1}{J_\theta},$$

where $\text{MSE}_\theta(\hat{\theta}_n^{\text{ML}})$ is the mean squared error of the estimator $\hat{\theta}_n^{\text{ML}}$. 

This equality holds for any estimator $\hat{\theta}_n$ that satisfies $\text{MSE}_\theta(\hat{\theta}_n) \geq \frac{1}{nJ_\theta}$.
when the family satisfies some regularity conditions**. Conversely, it is known that any suitable††, the estimator \( \hat{\theta}_n \) satisfies

\[
\text{nMSE}_\theta(\hat{\theta}_n) \gtrsim \frac{1}{J_\theta}
\]

when the number \( n \) of obtained data is large. Therefore, we can conclude that the inverse of Fisher information provides the optimal performance in the asymptotic setting.

### 6.3. Multi-Parametric Case

Next, we proceed to the multi-parametric case. We assume that \( n \) observations \( \tilde{\omega}_n \equiv (\omega_1, \ldots, \omega_n) \) are independently generated subject to the unknown probability distribution, which belongs to a multi-parametric probability family \( \{p_\theta|\theta = (\theta_1, \ldots, \theta_d) \in \Theta \subset \mathbb{R}^d\} \).

In this case, the estimator is denoted by a function \( \hat{\theta}_n \) from \( n \) observations to the parameter space \( \Theta \) or the larger set \( \mathbb{R}^d \), and we focus on the sum of MSE of respective parameters, or the MSE matrix:

\[
\text{MSE}^{i,j}_\theta(\hat{\theta}_n) \equiv \sum_{\tilde{\omega}_n} (\theta^i - \hat{\theta}^i_n(\tilde{\omega}_n))(\theta^j - \hat{\theta}^j_n(\tilde{\omega}_n))p_{\theta}^n(\tilde{\omega}_n),
\]

whose trace equals the sum of the MSEs of the respective parameters.

In this case, an estimator \( \hat{\theta}_n = (\hat{\theta}_n^1, \ldots, \hat{\theta}_n^d) \) is called unbiased if it satisfies

\[
\theta^i = \sum_{\tilde{\omega}_n} \hat{\theta}^i_n(\tilde{\omega}_n)p_{\theta}^n(\tilde{\omega}_n), \quad \forall \theta \in \Theta, \forall i.
\]

Any unbiased estimator \( \hat{\theta}_n \) satisfies the inequality

\[
(\text{MSE}^{i,j}_\theta(\hat{\theta}_n)) \geq \frac{1}{n}(J^{i,j}_\theta)^{-1},
\]

which is called the multi-parameter Cramér-Rao inequality, and \( (J^{i,j}_\theta) \) is the Fisher information matrix defined by

\[
J^{i,j}_\theta \equiv \sum_\omega \left( \frac{\partial \log p_\theta(\omega)}{\partial \theta^i} \right) \left( \frac{\partial \log p_\theta(\omega)}{\partial \theta^j} \right) p_\theta(\omega).
\]

Statistical researchers often focus on the MSE of the limiting distribution instead of the limit of MSE. This is because the inequalities (7) and (8) hold with less conditions in the version of limiting distributions [0-3]. However, there is a problem in a rigorous definition of such a ‘suitable estimator’. This problem is called superefficiency, and there are many different solutions in classical theory. For details, please refer to van der Vaart [0-3].
Furthermore, the maximum likelihood estimator (MLE) \( \hat{\theta}^n_{ML}(\bar{\omega}_n) \equiv \arg \max_{\theta} p^n_{\theta}(\bar{\omega}_n) \) satisfies
\[
\lim_{n \to \infty} n \text{MSE}^{i,j}_{\theta}(\hat{\theta}^n_{ML}) = (J^{i,j}_{\theta})^{-1},
\]
when the family satisfies some regularity conditions. Thus, regarding the sum of MSE, we have
\[
\lim_{n \to \infty} n \sum_{i=1}^{d} \text{MSE}^{i,i}_{\theta}(\hat{\theta}^n_{ML}) = \text{tr}((J^{i,j}_{\theta})^{-1}),
\]
where \((J^{i,j}_{\theta})^{-1}\) denotes the inverse of the matrix \((J^{i,j}_{\theta})\). Note that its trace \(\text{tr}((J^{i,j}_{\theta})^{-1})\) is different with \(\sum_{i=1}^{d} (J^{i,i}_{\theta})^{-1}\) in general.

Therefore, similarly to the one-parametric case, the maximum likelihood estimator is asymptotically optimal, and the inverse of Fisher information provides the optimal performance in the asymptotic setting. A similar relation of the multi-parametric case as (3) is summarized in Appendix K in Chap. 28.

Further Reading


CHAPTER 8
A New Approach to Cramér-Rao Bounds for Quantum State Estimation

Hiroshi Nagaoka

1. Introduction

Motivated mainly by the recent progress in the field of optical communications, the parameter estimation problem for a quantum statistical model \( S = \{ S_\theta; \theta = (\theta^1, \ldots, \theta^n) \in \Theta \} \), where \( \{ S_\theta \} \) are density operators on a Hilbert space, has been actively studied. The problem is to estimate the ‘true’ state which is unknown but is assumed to be represented by a density operator in \( \mathcal{H} \). This can be regarded as a noncommutative analogue of the parameter estimation problem for a ‘classical’ statistical model \( \mathcal{P} = \{ P_\theta; \theta = (\theta^1, \ldots, \theta^n) \in \Theta \} \), where \( \{ P_\theta \} \) are probability distributions, and many notions in the classical theory can find their analogues in the quantum theory. The unbiasedness of an estimator is one of these notions and is very important in the quantum case, also.

The fundamental theorem in the classical theory of unbiased estimation is the Cramér-Rao inequality which gives a lower bound for the variance of an arbitrary unbiased estimator. In the quantum case, however, several lower bounds of Cramér-Rao type (CR bounds, for short) are known. In particular, \( C^S \) which is based on the notion of ‘symmetric logarithmic derivatives’ [1, 2] and \( C^R \) which is based on the notion of ‘right logarithmic derivatives’ [5] are well-known. A. S. Holevo introduced and studied another CR bound \( C^H \) which, in a sense, unifies \( C^S \) and \( C^R \) [3, 4]. However, \( C^H \) is defined via some minimization procedure and is not generally written in an explicit form. Moreover, \( C^H \) is not ‘optimal’ in general, in the sense that there may be a CR bound which is more informative than \( C^H \). The theory of quantum CR bounds is not yet completed.

In this paper, some new results on quantum CR bounds are presented. First, we define the ‘most informative’ CR bound \( C^{MI} \) and derive a general
property of this bound. In the 1-parameter case \( n = 1 \), it is shown that \( C_{\text{MI}} = C_{\text{S}} \). Next, we introduce a new CR bound \( C_{\text{NEW}} \) for the 2-parameter case \( n = 2 \), which is based on a fundamental inequality on simultaneous measurements of two noncommuting observables. It is shown that \( C_{\text{MI}} \geq C_{\text{NEW}} \geq C_{\text{H}} \) in general. Furthermore, in the simplest noncommutative case \( \dim\mathcal{H} = 2 \), we show that \( C_{\text{MI}} = C_{\text{NEW}} \).

In this paper, we do not delve into the mathematically rigorous treatment of operators on a Hilbert space, but often proceed in an ‘intuitive’ manner, assuming several regularity conditions tacitly. For the rigorous argument, see [4].

2. Quantum Measurement Theory

In this section we give a brief review of the quantum measurement theory for the later discussion.

Suppose that a quantum physical system is given. Then a complex Hilbert space \( \mathcal{H} \) corresponds to the system, and physics of the system is described in terms of mathematics on \( \mathcal{H} \). We denote the inner product of two vectors \( \phi \) and \( \psi \) in \( \mathcal{H} \) by \( (\phi | \psi) \).

Let \( X \) be a linear operator on \( \mathcal{H} \). Then \( X^\ast \) (the adjoint operator of \( X \)) and \( \text{Tr} X \) (the trace of \( X \)) are defined as usual; i.e., \( (\phi | X \psi) = (X^\ast \phi | \psi) \) for \( \forall \phi, \forall \psi \) and \( \text{Tr} X = \sum_j (\psi_j | X \psi_j) \) where \( \{ \psi_j \} \) is an arbitrary orthonormal basis of \( \mathcal{H} \). It should be noted that \( \text{Tr} XY = \text{Tr} YX \) for \( \forall X, \forall Y \). When \( X = X^\ast \) holds, \( X \) is called a Hermitian operator. We write \( X > 0 (\geq 0) \) when \( (\psi | X \psi) > 0 (\geq 0) \) for \( \forall \psi \) nonzero.

A Hermitian operator \( S \) satisfying \( S \geq 0 \) and \( \text{Tr} S = 1 \) is called a density operator. It is also called a state, because a physical state of the quantum system is represented by a density operator. A state \( S \) is said to be pure if \( \text{rank} S = 1 \). In this case \( S \) becomes the orthogonal projection onto a 1-dimensional linear subspace of \( \mathcal{H} \). Using the Dirac’s notation, a pure state \( S \) is written as \( S = |\psi)(\psi| \) by a vector \( \psi \) of norm 1, i.e., \( (\psi|\psi) = 1 \). In general, any state \( S \) can be written as a mixture of pure states:

\[
S = \sum_j p_j |\psi_j)(\psi_j| \tag{1}
\]

where \( \{p_j\} \) are positive numbers satisfying \( \sum_j p_j = 1 \), and \( \{\psi_j\} \) are vectors of norm 1. Indeed, letting \( \{p_j\} \) be the nonzero eigenvalues of \( S \) and \( \{\psi_j\} \) be the corresponding eigenvectors constituting a orthonormal system, \( S \) is written as (1). A state which is not pure is called a mixed state.
Next we proceed to mathematical description of quantum measurements. Consider an arbitrary measurement with values in a set $\Omega$; i.e., if the measurement is performed to the quantum system, we get an element of $\Omega$ as the outcome. Owing to the statistical nature of quantum mechanics, the outcome fluctuates in general under a probability distribution on $(\Omega, \mathcal{B})$, where $\mathcal{B}$ is a $\sigma$-algebra consisting of subsets of $\Omega$. This probability distribution depends on the state of the system in the following way: when the present state of the system is represented by a density operator $S$, the probability that the outcome lies in $B$, where $B$ is a set in $\mathcal{B}$, is written as

$$P_S(B) = \text{Tr} SM(B)$$

where $M : B \mapsto M(B)$ is an operator-valued set function defined on $\mathcal{B}$ such that

(i) $M(\phi) = 0$, $M(\Omega) = I$ (the identity on $\mathcal{H}$);
(ii) $M(B) = M(B)^*$ $\geq 0$ for $\forall B \in \mathcal{B}$;
(iii) $M(\bigcup_j B_j) = \sum_j M(B_j)$ for any (at most countable) disjoint sequence $\{B_j\}$ of sets in $\mathcal{B}$.

It should be noted that $M$ is determined by the considered measurement and does not depend on the state $S$. Motivated by this fact, a measurement on a measurable space $(\Omega, \mathcal{B})$ is mathematically defined as an operator-valued set function $M$ satisfying (i)–(iii) above. (Such an $M$ is also called a probability operator-valued measure on $(\Omega, \mathcal{B})$.) We often omit $\mathcal{B}$ from $(\Omega, \mathcal{B})$ when $\mathcal{B}$ is the ‘natural’ $\sigma$-algebra on $\Omega$ such as the totality of Borel subsets in the case of a topological $\Omega$ or the totality of subsets in the case of finite $\Omega$.

If a measurement $M$ on $(\Omega, \mathcal{B})$ is projection-valued in the sense that

(iv) $M(B)^2 = M(B)$ for $\forall B \in \mathcal{B}$,

we say that $M$ is simple. It can be shown that, under the conditions (i)–(iii), (iv) is equivalent to the orthogonality condition:

(v) $M(B_1)M(B_2) = 0$ if $B_1 \cap B_2 = \phi$.

In many standard texts of quantum mechanics, only this type of measurements is referred to as the representation of measurements. It is known that an arbitrary measurement over the system $\mathcal{H}$ can be ‘realized’ by a simple measurement over the ‘composite system’ of $\mathcal{H}$ and an additional independent system $\mathcal{H}_0$ in a fixed state $S_0$, (see [4]).
Simple measurements on $\mathbb{R}$ (the set of all real numbers) are of particular importance because they represent the measurements of real-valued physical quantities called observables. As is well-known, an arbitrary observable corresponds to a Hermitian operator, say $X$, which has the spectral representation

$$X = \int_{-\infty}^{\infty} x E(dx),$$

(2)

where $E$ is a simple measurement on $\mathbb{R}$. This $E$ represents the measurement of the observable corresponding to $X$. Note that the correspondence between Hermitian operators and simple measurements on $\mathbb{R}$ by (2) is one-to-one.

3. Parameter Estimation and Lower Bounds of Cramér-Rao Type (CR Bounds)

Let $\Theta$ be an open set in $\mathbb{R}^n$, and let $S = \{S_{\theta}; \theta \in \Theta\}$ be a family of density operators on a Hilbert space $\mathcal{H}$ smoothly parametrized by an $n$-dimensional parameter $\theta = (\theta^1, \ldots, \theta^n)$ with range $\Theta$. Such a family is called an $n$-dimensional quantum statistical model. We consider the parameter estimation problem for the model $S$.

An estimator for the parameter $\theta$ based on some measurement is itself regarded as a measurement with outcomes in the parameter space $\Theta$, and is represented by a probability operator-valued measure on $\Theta$. For a technical reason, we define an estimator as a measurement on $\mathbb{R}^n$, permitting the possibility of getting an estimate out of $\Theta$. Such a definition is often adopted in the classical statistical theory, also.

Let $\hat{M}$ be an estimator. When the system is in the state $S_{\theta}$ for some $\theta \in \Theta$, the probability that the estimate lies in a Borel set $B$ in $\mathbb{R}^n$ is written as

$$\hat{P}_{\theta}(B) = \text{Tr} S_{\theta} \hat{M}(B).$$

$\hat{M}$ is said to be unbiased if for $\forall \theta \in \Theta$

$$\int \hat{\theta} \hat{P}_{\theta}(d\theta) = \theta^j \quad (j = 1, 2, \ldots, n).$$

(3)

Differentiating this equation we obtain

$$\int \hat{\theta} \frac{\partial}{\partial \theta^k} \hat{P}_{\theta}(d\theta) = \delta^j_k \quad (j, k = 1, 2, \ldots, n),$$

(4)
where \( \delta^i_j \) is the Kronecker’s delta. When \( \hat{M} \) satisfies (3) and (4) at a fixed point \( \theta \in \Theta \), we say that \( \hat{M} \) is \textit{locally unbiased at} \( \theta \). Obviously, an estimator is unbiased if and only if it is locally unbiased at every \( \theta \in \Theta \).

Suppose that \( \hat{M} \) is locally unbiased at a fixed point \( \theta \). Then the covariance matrix of \( \hat{M} \) at the state \( S_\theta \), say \( V_\theta[\hat{M}] = [v^{jk}_\theta] \in \mathbb{R}^{n \times n} \), is defined as

\[
v^{jk}_\theta = \int (\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k) \hat{P}_\theta(d\hat{\theta}).
\]

We describe the accuracy of the estimation at \( \theta \) by

\[
\text{tr} G V_\theta[\hat{M}] = \sum_{j,k} g_{jk} v^{jk}_\theta,
\]

where \( \text{tr} \) denotes the trace of an \( n \times n \) matrix in distinction from \( \text{Tr} \) for an operator on \( \mathcal{H} \), and \( G = [g_{jk}] \in \mathbb{R}^{n \times n} \) is a positive-definite matrix.

Now we define a \textit{lower bound of Cramér-Rao type} (CR bound, for short) (with respect to \( G \) at \( \theta \)) as a quantity \( C \) which satisfies

\[
\text{tr} G V_\theta[\hat{M}] \geq C
\]

for any estimator \( \hat{M} \) which is locally unbiased at \( \theta \).

From now on, we assume that any state \( S_\theta \) in \( \mathcal{S} \) is nondegenerate; i.e., \( S_\theta > 0 \) for \( \forall \theta \in \Theta \). This assumption on the model \( \mathcal{S} \) will simplify the argument on CR bounds very much.

\section{Some Examples of CR Bounds}

We introduce two well-known CR bounds here. Suppose that \( G \) and \( \theta \) are given and fixed. First, as quantum analogues of

\[
\frac{\partial}{\partial \theta^j} \log p_\theta = \frac{1}{p_\theta} \frac{\partial}{\partial \theta^j} p_\theta
\]

for a (classical) statistical model of probability distribution functions \( \{p_\theta\} \), we define the \textit{symmetric logarithmic derivatives} at \( \theta \) \( \{L^S_{\theta,j}; j = 1, \ldots, n\} \) and the \textit{right logarithmic derivatives} at \( \theta \) \( \{L^R_{\theta,j}; j = 1, \ldots, n\} \) as follows: \( L^S_{\theta,j} \) is a Hermitian operator on \( \mathcal{H} \) such that

\[
\frac{\partial}{\partial \theta^j} S_\theta = \frac{1}{2} (S_\theta L^S_{\theta,j} + L^S_{\theta,j} S_\theta),
\]

and \( L^R_{\theta,j} \) is a linear operator on \( \mathcal{H} \) such that

\[
\frac{\partial}{\partial \theta^j} S_\theta = S_\theta L^R_{\theta,j}.
\]
Using these logarithmic derivatives, quantum analogues of the Fisher information matrix are defined: \( J^S_\theta \) be an \( n \times n \) real symmetric matrix whose \((j,k)\) element is

\[
[J^S_\theta]_{jk} = \operatorname{Re} \operatorname{Tr} S_\theta L^S_{\theta,j} L^S_{\theta,k},
\]

and \( J^R_\theta \) be an \( n \times n \) complex Hermitian matrix whose \((j,k)\) element is

\[
[J^R_\theta]_{jk} = \operatorname{Tr} S_\theta L^R_{\theta,k} (L^R_{\theta,j})^*.
\]

Let

\[
C^S \overset{\text{def}}{=} \operatorname{tr} G(J^S_\theta)^{-1}, \quad C^R \overset{\text{def}}{=} \operatorname{tr} \operatorname{Re}(J^R_\theta)^{-1} + \operatorname{tr} \operatorname{abs} G \operatorname{Im}(J^R_\theta)^{-1},
\]

where \( \operatorname{abs} A \) for a matrix \( A \) is defined by

\[
\operatorname{abs} A \overset{\text{def}}{=} T \begin{bmatrix} |\alpha_1| & & \\ & \ddots & \\ & & |\alpha_n| \end{bmatrix} T^{-1}
\]

when \( A \) is diagonalized as

\[
A = T \begin{bmatrix} \alpha_1 \\ & \ddots \\ & & \alpha_n \end{bmatrix} T^{-1}
\]

and therefore we have: \( \operatorname{tr} \operatorname{abs} A = \sum_j |\alpha_j| \). Then it can be shown that both \( C^S \) and \( C^R \) are CR bounds \([1, 2, 4, 5]\).

A.S.Holevo introduced another CR bound, say \( C^H \), through which we can understand \( C^S \) and \( C^R \) in a unified manner \([3, 4]\). \( C^H \) is based on the following lemma.

**Lemma 1:** Suppose that

\[
\int_\Omega f(\omega) M(d\omega) = F
\]

where \( M \) is a measurement on a set \( \Omega \), \( f \) is a complex-valued function defined on \( \Omega \), and \( F \) is an operator on \( \mathcal{H} \). Then we have

\[
\int_\Omega |f(\omega)|^2 M(d\omega) \geq FF^*.
\]

**Proof:**

\[
\int_\Omega |f(\omega)|^2 M(d\omega) - FF^* = \int_\Omega \{f(\omega) - F\} M(d\omega)\{f(\omega) - F\}^* \geq 0.
\]
Let \( \hat{M} \) be an estimator which is locally unbiased at \( \theta \), and let
\[
X^j \overset{\text{def}}{=} \int (\hat{\theta}^j - \theta^j) \hat{M}(d\hat{\theta}) \quad (j = 1, 2, \ldots, n).
\] (5)

It is noted that the local unbiasedness condition (3)(4) is written as
\[
\text{Tr} \, S_{\theta} X^j = 0 \quad (j = 1, 2, \ldots, n); \quad (6)
\]
\[
\text{Tr} \, \frac{\partial}{\partial \theta^k} S_{\theta} X^j = \delta^j_k \quad (j, k = 1, 2, \ldots, n). \quad (7)
\]

For an arbitrary complex vector \( \xi = [\xi^j] \in \mathbb{C}^n \), define the function \( f : \mathbb{R}^n \to \mathbb{C} \) and the operator \( F \) by
\[
f(\hat{\theta}) = \sum_{j=1}^{n} \xi_j (\hat{\theta}^j - \theta^j), \quad F = \sum_{j=1}^{n} \xi_j X^j.
\]

Then we have \( \int f(\hat{\theta}) \hat{M}(d\hat{\theta}) = F \) from (5), and therefore Lemma 1 claims that
\[
\int |f(\hat{\theta})|^2 \hat{M}(d\hat{\theta}) \geq FF^*,
\]
from which we obtain
\[
\int |f(\hat{\theta})|^2 \text{Tr} \, S_{\theta} \hat{M}(d\hat{\theta}) \geq \text{Tr} \, S_{\theta} FF^*.
\]

The left-hand side of this inequality is written as \( \xi^* V_{\theta}[\hat{M}] \xi \) by means of the covariance matrix \( V_{\theta}[\hat{M}] \), while the right-hand side is written as \( \xi^* Z_{\theta}[\vec{X}] \xi \) (where \( \vec{X} = (X^1, \ldots, X^n) \)) by means of \( Z_{\theta}[\vec{X}] = [z^{jk}] \in \mathbb{C}^{n \times n} \) such that
\[
z^{jk} = \text{Tr} \, S_{\theta} X^k X^j. \quad (8)
\]

Thus we have
\[
V_{\theta}[\hat{M}] \geq Z_{\theta}[\vec{X}].
\]

The above consideration leads us to define the CR bound \( C^H \) as
\[
C^H \overset{\text{def}}{=} \min \{ \lambda_{\theta}(\vec{X}; G) \mid \vec{X} \in \mathcal{X}_\theta \}
\]
where
\[
\lambda_{\theta}(\vec{X}; G) \overset{\text{def}}{=} \min \left\{ \text{tr} \, GV \mid V \text{ is an } n \times n \text{ real symmetric matrix satisfying } V \geq Z_{\theta}[\vec{X}] \right\}
\]
\[
= \text{tr} \, G \text{ Re} Z_{\theta}[\vec{X}] + \text{tr} \, \text{abs} \, G \text{ Im} Z_{\theta}[\vec{X}], \quad ([4])
\]
\[ \mathcal{X}_\theta \overset{\text{def}}{=} \{ \vec{X} = (X^1, \ldots, X^n) \mid \{X^j\} \text{ are Hermitian operators satisfying (6) and (7)} \}. \]

\( C^H \) is ‘more informative’ than both \( C^S \) and \( C^R \) in the sense that \( C^H \geq \max\{C^S, C^R\} \). This can be seen immediately from the following relations:

\[ C^S = \min \{ \text{tr} G \text{Re} Z_\theta[\vec{X}] \mid \vec{X} \in \mathcal{X}_\theta \}, \quad C^R = \min \{ \lambda_\theta(\vec{X}; G) \mid \vec{X} \in \tilde{\mathcal{X}}_\theta \} \]

where \( \tilde{\mathcal{X}}_\theta \) is an extension of \( \mathcal{X}_\theta \) such that

\[ \tilde{\mathcal{X}}_\theta \overset{\text{def}}{=} \{ \vec{X} = (X^1, \ldots, X^n) \mid \{X^j\} \text{ are linear operators and satisfy (6) and (7)} \} \]

and, correspondingly, \( \lambda_\theta(\vec{X}; G) \) is extended to be defined for non-Hermitian \( \vec{X} \) by rewriting (8) to

\[ z_{jk} = \text{Tr} S_\theta X^k(X^j)^*. \]

5. The Most Informative CR Bound

The most informative CR bound \( C^{\text{MI}} \) is defined as the maximum of all the CR bounds, or equivalently

\[ C^{\text{MI}} \overset{\text{def}}{=} \min \{ \text{tr} G V_\theta[\hat{M}] \mid \hat{M} \text{ is locally unbiased at } \theta \}. \]

It is also written as

\[ C^{\text{MI}} = \min \{ \nu_\theta(\vec{X}; G) \mid \vec{X} \in \mathcal{X}_\theta \} \quad (9) \]

where

\[ \nu_\theta(\vec{X}; G) \overset{\text{def}}{=} \min \{ \text{tr} G V_\theta[\hat{M}] \mid \hat{M} \text{ is an estimator satisfying} \]

\[ X^j = \int (\hat{\theta}^j - \theta^j) \hat{M}(d\hat{\theta}), \quad j = 1, \ldots, n \}. \]

So, the study of \( C^{\text{MI}} \) can be divided into two steps: the first step is to seek for an explicit expression of \( \nu_\theta(\vec{X}; G) \), and the second step is to solve the minimization problem (9). Actually, we will carry out this program for a special case later.

Now we elucidate the relation between \( C^{\text{MI}} \) and the Cramér-Rao bound of the classical statistical model obtained from \( S \) by fixing a measurement. Let \( M \) be a measurement on \( (\Omega, \mathcal{B}) \) and let

\[ P^M_\theta(B) \overset{\text{def}}{=} \text{Tr} S_\theta M(B) \quad (B \in \mathcal{B}, \ \theta \in \Theta). \]
Then $\mathcal{P}^M \overset{\text{def}}{=} \{ P^M_\theta ; \theta \in \Theta \}$ becomes a classical statistical model on $(\Omega, \mathcal{B})$. Since $\{ S_\theta \}$ are assumed to be nondegenerate, the condition $P^M_\theta(B) = 0$ is equivalent to $M(B) = 0$ and hence does not depend on $\theta$. This means that there exists a measure $\mu$ on $(\Omega, \mathcal{B})$ such that, for $\forall \theta \in \Theta$, the density function of $P^M_\theta$ w.r.t. $\mu$, say $p^M_\theta = dP^M_\theta / d\mu$, is defined and is positive almost everywhere. So, we can define the Fisher information matrix $J^M_\theta \in \mathbb{R}^{n \times n}$ as

$$
[J^M_\theta]_{jk} \overset{\text{def}}{=} \int \left\{ \frac{\partial}{\partial \theta^j} \log p^M_\theta(\omega) \right\} \left\{ \frac{\partial}{\partial \theta^k} \log p^M_\theta(\omega) \right\} P^M_\theta(d\omega).
$$

**Theorem 2:**

$$C^{\text{MI}} = \min \{ \text{tr} G(J^M_\theta)^{-1} \mid M \text{ is a measurement} \}.$$  

**Proof:** Denote the right-hand side by $C$. The (classical) Cramér-Rao theorem says that $V_\theta[\hat{M}] \geq (J^M_\theta)^{-1}$ for an arbitrary estimator $\hat{M}$ which is locally unbiased at $\theta$. This leads to $C^{\text{MI}} \geq C$. On the other hand, for an arbitrary measurement $M$ on a set $\Omega$, define the mapping $\hat{\theta}: \Omega \rightarrow \mathbb{R}^n$ by

$$\hat{\theta}^j(\omega) = \theta^j + \sum_{k=1}^n [(J^M_\theta)^{-1}]^{jk} \frac{\partial}{\partial \theta^k} \log p^M_\theta(\omega), \quad (j = 1, 2, \ldots, n, \ \omega \in \Omega),$$

and define the estimator $\hat{M}$ by

$$\hat{M}(B) = M(\hat{\theta}^{-1}(B)) \quad (B: \text{a Borel set in } \mathbb{R}^n).$$

It then turns out that $\hat{M}$ is locally unbiased at $\theta$ and satisfies $V_\theta[\hat{M}] = (J^M_\theta)^{-1}$. This leads to $C^{\text{MI}} \leq C$. \hfill \blacksquare

Here we consider the single parameter case $n = 1$. Let $E_\theta$ be the spectral resolution of the Hermitian operator $\theta + (J^S_\theta)^{-1} L^S_\theta$, i.e., $E_\theta$ is a simple measurement on $\mathbb{R}$ such that

$$\theta + (J^S_\theta)^{-1} L^S_\theta = \int \hat{\theta} E_\theta(d\hat{\theta}).$$

Then it follows from the definition of $L^S_\theta$ that $E_\theta$ is locally unbiased at $\theta$. Moreover, noting that $E_\theta$ is a simple measurement, we can see that $V_\theta[E_\theta] = (J^S_\theta)^{-1}$. This leads to

$$C^{\text{MI}} = C^S,$$

and from Theorem 2 we have

$$J^S_\theta = J^{E_\theta}_\theta = \max \{ J^M_\theta \mid M \text{ is a measurement} \}.$$  

This means that $E_\theta$ is most informative among all the measurements with respect to the parameter estimation problem for $\mathcal{S}$ at the fixed point $\theta$. 

6. A New CR Bound for the 2-Parameter Case

In this section we introduce a new CR bound $C^\text{NEW}$ for the two parameter case $n = 2$. The following lemma is fundamental.

**Lemma 3:** Let $A$ and $B$ be Hermitian operators and let $S$ be a nondegenerate density operator. If a measurement $M$ on $\mathbb{R}^2$ satisfies

$$\int aM(da \, db) = A, \quad \int bM(da \, db) = B \quad (10)$$

then

$$\int (a^2 + b^2)\mathrm{Tr} SM(da \, db) \geq \mathrm{Tr} S(A^2 + B^2) + \mathrm{Tr} \mathrm{Abs} S[A, B] \quad (11)$$

where $[A, B] \overset{\text{def}}{=} AB - BA$ and, for an operator $X$ having the eigenvalues $\{\lambda_j\}$, $\mathrm{Tr} \mathrm{Abs} X \overset{\text{def}}{=} \sum_j |\lambda_j|$. 

**Proof:** Since (10) is written as

$$\int (a \pm ib)M(da \, db) = A \pm iB, \quad (i^2 = -1)$$

it follows from Lemma 1 that

$$\int (a^2 + b^2)M(da \, db) = \int |a \mp ib|^2 M(da \, db) \geq (A \mp iB)(A \pm iB) = A^2 + B^2 \pm i[A, B],$$

which leads to

$$\int (a^2 + b^2)S^{\frac{1}{2}}M(da \, db)S^{\frac{1}{2}} \geq S^{\frac{1}{2}}(A^2 + B^2)S^{\frac{1}{2}} \pm iS^{\frac{1}{2}}[A, B]S^{\frac{1}{2}}. \quad (12)$$

Let $\{\lambda_j\}$ be the eigenvalues of the Hermitian operator $iS^{\frac{1}{2}}[A, B]S^{\frac{1}{2}}$ and $\{\psi_j\}$ be the corresponding eigenvectors constituting an orthonormal basis. Then from (12) we have

$$\int (a^2 + b^2)(\psi_j | S^{\frac{1}{2}}M(da \, db)S^{\frac{1}{2}}\psi_j) \geq (\psi_j | S^{\frac{1}{2}}(A^2 + B^2)S^{\frac{1}{2}}\psi_j) + |\lambda_j|.$$

Applying $\sum_j$ to the both sides, we obtain

$$\int (a^2 + b^2)\mathrm{Tr} SM(da \, db) \geq \mathrm{Tr} S(A^2 + B^2) + \mathrm{Tr} \mathrm{Abs} S^{\frac{1}{2}}[A, B]S^{\frac{1}{2}},$$

which is identical with (11) because the eigenvalues of $S^{\frac{1}{2}}[A, B]S^{\frac{1}{2}}$ are equal to those of $S[A, B] = S^{\frac{1}{2}}(S^{\frac{1}{2}}[A, B]S^{\frac{1}{2}})S^{-\frac{1}{2}}$. 

\[\blacksquare\]
Remark 4: In the above lemma we assumed that \( iS^{1/2}[A, B]S^{1/2} \) has discrete eigenvalues. Some mathematical refinement would be desired so that the lemma is reformulated and is proved in a more general setting.

Remark 5: It is well-known that if \([A, B] = 0\) there exists a simple measurement \( M \) satisfying (10). In this case, we have \( \int (a^2 + b^2)M(da\,db) = A^2 + B^2 \), and therefore the equality in (11) holds for \( \forall S \). Next, suppose that \( A \) and \( B \) satisfy the Heisenberg canonical commutation relation \([A, B] = i\hbar \) (\( \hbar \) is a positive constant). Then, for \( \forall a \in \mathbb{R}, \forall b \in \mathbb{R} \), we can define \( \psi = |a, b\rangle \) as a vector of norm 1 satisfying \((A + iB)\psi = (a + ib)\psi\), and a measurement \( M \) is defined by

\[
M(da\,db) \overset{\text{def}}{=} |a, b\rangle(b, a)|da\,db|^{2\pi\hbar}.
\]

This measurement is shown to satisfy (10) and \( \int (a^2 + b^2)M(da\,db) = A^2 + B^2 + \hbar \) ([4]). Again, the equality in (11) holds for \( \forall S \). For arbitrarily given \( A \) and \( B \), however, there does not always exist a measurement \( M \) satisfying (10) and achieving the lower bound in (11) for \( \forall S \) simultaneously. Nevertheless, we can consider the problem whether, for arbitrarily given \( A, B \) and \( S \), there always exists a measurement \( M \) satisfying (10) and achieving the lower bound in (11). This has been affirmatively solved for the case \( \dim \mathcal{H} = 2 \) by the author*, but is still an open problem for the general case at present.

Now, suppose that we are given a 2-parameter model \( S = \{S_\theta; \theta = (\theta^1, \theta^2) \in \Theta\} \) and a positive-definite matrix \( G = [g_{jk}] \in \mathbb{R}^{2 \times 2} \). Let

\[
C^{\text{NEW}} \overset{\text{def}}{=} \min \{\mu_\theta(\bar{X}; G) | \bar{X} = (X^1, X^2) \in \mathcal{X}_\theta \}
\]

where

\[
\mu_\theta(\bar{X}; G) \overset{\text{def}}{=} \text{tr} G \text{Re}Z_\theta[\bar{X}] + \sqrt{\text{det} G} \text{Tr} \text{Abs} S_\theta[X^1, X^2].
\]

We have the following theorem.

**Theorem 6:** \( C^{\text{NEW}} \) is a CR bound, and is more informative than \( C^H \).

**Proof:** Let \( \hat{M} \) be an arbitrary estimator, and let

\[
a(\hat{\theta}) \overset{\text{def}}{=} \sum_{k=1}^{2} h_{1k}(\hat{\theta}^k - \theta^k), \quad b(\hat{\theta}) \overset{\text{def}}{=} \sum_{k=1}^{2} h_{2k}(\hat{\theta}^k - \theta^k)
\]

*Additional author’s note: See Chap. 11 of this book.*
where \( h_{jk} \) denotes the \((j,k)\) element of \( G^{\frac{1}{2}} \). Then we have

\[
\text{tr} \, G V_\theta [\hat{M}] = \int \left[ \{ a(\hat{\theta}) \}^2 + \{ b(\hat{\theta}) \}^2 \right] \text{Tr} \, S_\theta \hat{M} (d\hat{\theta}).
\]

Hence, defining

\[
A \overset{\text{def}}{=} \int a(\hat{\theta}) \hat{M} (d\hat{\theta}) = \sum_{k=1}^{2} h_{1k} X^k, \quad B \overset{\text{def}}{=} \int b(\hat{\theta}) \hat{M} (d\hat{\theta}) = \sum_{k=1}^{2} h_{2k} X^k,
\]

where \((X^1, X^2)\) are defined from \( \hat{M} \) by (5), we obtain from Lemma 3

\[
\text{tr} \, G V_\theta [\hat{M}] \geq \text{Tr} \, S_\theta \left[ A^2 + B^2 \right] + \text{Tr} \, \text{Abs} S_\theta [A, B] = \mu_\theta (\vec{X}; G).
\]

This inequality immediately proves the first assertion of the theorem that \( C^\text{NEW} \) is a CR bound.

The second assertion that \( C^\text{NEW} \geq C^\text{H} \) follows from the inequality

\[
\mu_\theta (\vec{X}; G) \geq \lambda_\theta (\vec{X}; G) \quad (\forall \vec{X} = (X^1, X^2)),
\]

which is proved in the following way. It is seen that

\[
G \text{Im} Z_\theta [\vec{X}] = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} 0 & \frac{i}{2} \text{Tr} S_\theta [X^1, X^2] \\ -\frac{i}{2} \text{Tr} S_\theta [X^1, X^2] & 0 \end{bmatrix}
\]

\[
= \frac{i}{2} \text{Tr} S_\theta [X^1, X^2] \begin{bmatrix} -g_{12} & g_{11} \\ -g_{22} & g_{21} \end{bmatrix}.
\]

This matrix has the eigenvalues \( \pm \frac{1}{2} \sqrt{\det G} \, \text{Tr} S_\theta [X^1, X^2] \), and hence we have

\[
\text{tr} \, G \text{Im} Z_\theta [\vec{X}] = \sqrt{\det G} \, \left| \text{Tr} S_\theta [X^1, X^2] \right| \leq \sqrt{\det G} \, \text{Tr} \, \text{Abs} S_\theta [X^1, X^2],
\]

which yields (13).

In the simplest noncommutative case \( n = 2 \), which is known to describe the spin of a spin-\( \frac{1}{2} \) particle, we have the following very strong theorem. Regrettably, the proof is omitted here for want of space.

**Theorem 7:** If \( \dim \mathcal{H} = 2 \), we have

\[
C^\text{NEW} = C^\text{MI} = C^S + \left( \sqrt{\det G} / \det \, J_\theta^S \right) \text{Tr} \, \text{Abs} S_\theta [L_{\theta,1}^S, L_{\theta,2}^S].
\]
References


Additional Note by Author

Invoking some results from the paper “A generalization of the simultaneous diagonalization of Hermitian matrices and its relation to quantum estimation theory” included in this volume, Theorem 7 is readily proved as follows. For a pair of Hermitian operators \( \vec{X} = (X_1, X_2) \), let

\[
\Phi[\vec{X}] \doteq \text{Re} Z_\theta[\vec{X}] = \begin{bmatrix}
    \langle X_1, X_1 \rangle & \langle X_1, X_2 \rangle \\
    \langle X_2, X_1 \rangle & \langle X_2, X_2 \rangle
\end{bmatrix},
\]

where \( \langle X, Y \rangle \doteq \text{ReTr} S_\theta XY \). It then follows from Theorem 2 and Lemma 8 in “A generalization of ...” that

\[
\mu_\theta(\vec{X}; G) = \nu_\theta(\vec{X}; G) = \text{tr} G\Phi[\vec{X}] + 2\sqrt{\text{det} G\Phi[\vec{X}]}.
\]

In addition, letting \( \vec{K} = (K_1, K_2) \) be defined by

\[
K_j = \sum_{k=1}^{2} [(J_{\theta}^S)^{-1}]_{jk} L_{\theta, k}^S,
\]

we have \( \vec{K} \in \mathcal{X}_\theta \) and \( \Phi[\vec{X}] \geq \Phi[\vec{K}] = (J_{\theta}^S)^{-1} \) for any \( \vec{X} \in \mathcal{X}_\theta \), which is the essence of the fact that \( C^S \) is a CR bound. Therefore we obtain

\[
C^{\text{NEW}} = C^{\text{MI}} = \min_{\vec{X} \in \mathcal{X}_\theta} \left( \text{tr} G\Phi[\vec{X}] + 2\sqrt{\text{det} G\Phi[\vec{X}]} \right)
= \text{tr} G(J_{\theta}^S)^{-1} + 2\sqrt{\text{det} G(J_{\theta}^S)^{-1}}.
\]

The expression of Theorem 7 is obtained from this by applying Lemma 8 in “A generalization of ...” again to yield \( \text{Tr Abs} S_\theta[L_{\theta, 1}^S, L_{\theta, 2}^S] = 2\sqrt{\text{det} J_{\theta}^S} \).
CHAPTER 9

On Fisher Information of Quantum Statistical Models

Hiroshi Nagaoka

1. Introduction

In various fields of up-to-date sciences and technologies, quantum noise (quantum mechanical fluctuation) is a big obstacle (one is beyond far and one is impending immediately) in the way of their progress. The characteristic of quantum noise is specified by the quantum mechanical state (statistical operator) of the system, which is considerably different from the way that the characteristic of the ordinary noise is specified by the probability distribution. Therefore, classical statistics and information theory based on the ordinary probability theory are not sufficient to deal with the quantum noise, and we newly need to develop quantum statistics and quantum information theory based on quantum mechanical probability theory (that is also called non-commutative probability theory). This kind of theory is mainly studied in relation to optical communications now, and it is expected that it will be applied to various fields by and by. Moreover, the results of such theoretical researches may also suggest new technical feasibilities in the future.

In this paper, we will consider a parametric estimation problem with respect to a one-parameter model \( \{P_\theta \mid \theta \in \Theta\} \) (\( \Theta \subset \mathbb{R} \)) consisting of statistical operators. This is the quantum version of an ordinary (classical) parametric estimation problem. In the classical case, the accuracy of estimation can be evaluated by the so-called Fisher information \( I(\theta) \). That is, the variance of an unbiased estimator cannot be smaller than \( 1/I(\theta) \) (Cramér-Rao inequality). If there exists an unbiased estimator attaining this lower bound, the model belongs to the class called exponential family, and it is known to have various good properties. In addition, even in a general model, this lower bound is attainable asymptotically (as data size \( \to \infty \)).

This paper was originally written in Japanese. It was translated to English by Fuyuhiko Tanaka and Yoshiyuki Tsuda.
Moreover, through consideration on the independence and the sufficiency of a statistic, \( I(\theta) \) is provided with a meaning as the *information of data with respect to estimation of \( \theta \). On the other hand, in the case of the quantum system, analogues of the Fisher information are defined based on symmetric logarithmic derivatives or right logarithmic derivatives. It is shown that *quantum Cramér-Rao inequalities* hold with respect to them [1, 2, 3]. However, the definition of them gives an impression as if they were defined ad hoc so that the Cramér-Rao inequalities hold. Moreover, the meaning of them as information is not obvious. In this paper, we will investigate the Fisher information defined by the symmetric logarithmic derivatives from various points of view and consider its meaning.

In the sequel, the Hilbert space of the quantum system of interest is assumed to be finite-dimensional. Although this assumption causes a considerable loss of generality, the essence of the following arguments should be valid even in the infinite-dimensional case. Due to this assumption, we are free from complications accompanying functional-analytical arguments and are able to argue quantum mechanical non-commutativity in the framework of elementary linear algebra.

### 2. Statistical Operator and Non-Commutative Probability Theory

In this section, we explain the basic idea of non-commutative probability theory. Now, suppose that the quantum system of interest is described by a finite-dimensional complex Hilbert space \( \mathcal{H} \). Let the inner product on \( \mathcal{H} \) be denoted by \((x, y) \in \mathbb{C} \ (x, y \in \mathcal{H})\), and let \( \mathcal{L} \) be the set of linear operators (linear transformations) on \( \mathcal{H} \). Then, for each \( A \in \mathcal{L} \), the conjugate operator \( A^* \in \mathcal{L} \) is defined by

\[
(x, Ay) = (A^*x, y) \quad (\forall x, \forall y \in \mathcal{H}).
\]

An operator \( A \in \mathcal{L} \) is called a Hermitian operator iff \( A = A^* \) holds, and the set of Hermitian operators is denoted by \( \mathcal{L}_h \). Each \( A \in \mathcal{L}_h \) corresponds to an observable in the quantum system \( \mathcal{H} \), and the eigenvalues of \( A \) represent values which the observable can take. A Hermitian operator plays a role of a random variable in the ordinary probability theory.

One can perform various measurements on the quantum system \( \mathcal{H} \), and each measurement corresponds in a one-to-one manner to a family of oper-
ators $\Pi = \{\pi_k \mid k = 1, 2, \ldots\}$ satisfying the following:

\[
\begin{align*}
\pi_k &= \pi_k^* \in \mathcal{L}_h, \\
\pi_k^2 &= \pi_k, \\
k \neq l &\Rightarrow \pi_k \pi_l = 0, \\
\sum_k \pi_k &= \text{Id} \quad \text{(identity)}.
\end{align*}
\]

Moreover, for the measurement corresponding to $\Pi$, there is a one-to-one correspondence between an outcome of the measurement and a $\pi_k$. The conditions (1) and (2) are equivalent to that $\pi_k$ is the orthogonal projection to a subspace $V_k$ of $\mathcal{H}$, and the conditions (3) and (4) imply that the $\{V_k\}$ satisfies

\[
k \neq l \Rightarrow V_k \perp V_l \quad \text{(orthogonal)}, \quad \bigoplus_k V_k = \mathcal{H}.
\]

A family of operators $\Pi = \{\pi_k\}$ satisfying (1)-(4) is called an \textit{orthogonal resolution of identity} or a \textit{measurement} \textsuperscript{*}. (Note that, in this paper, we do not consider the notion of \textit{generalized quantum measurement} \textsuperscript{[1, 2, 3]}.)

Any observable $A \in \mathcal{L}_h$ can be written as

\[
A = \sum_k a_k \pi_k
\]

by a measurement $\Pi = \{\pi_k\}$. For example, we can choose the orthogonal projection to the eigenspace $V_k = \{x \mid Ax = a_k x\}$ as $\pi_k$. In this case,

\[
k \neq l \Rightarrow a_k \neq a_l
\]

holds, and then (5) is called the \textit{spectral decomposition} of $A$. In general, equation (5) is interpreted to mean that the value of the observable $A$ is determined by the measurement $\Pi$. Namely, when the outcome of the measurement corresponding to $\pi_k$ is obtained, the value of $A$ is determined to be $a_k$.

The quantum system $\mathcal{H}$ can take various \textit{states}. Each state corresponds in a one-to-one manner to a $P \in \mathcal{L}_h$ satisfying the following:

\[
\begin{align*}
P &= P^* \geq 0 \quad \text{(i.e. } (x, Px) \geq 0 \quad \forall x \in \mathcal{H}), \\
\text{Tr } P &= 1,
\end{align*}
\]

\textsuperscript{*}Additional author’s note: This is another name of the POVM.
where, in general, Tr $A$ ($A \in \mathcal{L}$) denotes the trace of $A$, and, by using an arbitrary orthonormal basis $\{u_1, u_2, \ldots\}$ of $\mathcal{H}$, it can be written as

$$\text{Tr} \ A = \sum_k (u_k, Au_k).$$

Note that Tr $AB = \text{Tr} BA$. An operator $P$ satisfying (7) and (8) is called a statistical operator, a density operator, a state etc., and it plays a role of a probability distribution in the ordinary probability theory. When we perform the measurement $\Pi$ under the state $P$, the probability of obtaining the outcome $\pi_k$, say $p(k \mid \Pi)$, is given by

$$p(k \mid \Pi) = \text{Tr} \ P\pi_k.$$ (9)

In terms of the spectral decomposition (5) and (6) of the observable $A$, equation (9) represents the probability of obtaining $a_k$ as the outcome of the measurement of $A$. From this, the expected value of $A$ is given by the following formula:

$$E_P[A] = \sum_k a_k \text{Tr} \ P\pi_k = \text{Tr} \ PA.$$ (10)

Hereafter, we focus on the regular statistical operators $P$ (so that $P > 0$) and denote the set of them by $\mathcal{P}$. Moreover, for each $P \in \mathcal{P}$, a complex inner product $(\cdot, \cdot)_P$ and a real inner product $\langle \cdot, \cdot \rangle_P$ on $\mathcal{L}$ are defined by the following:

$$(A, B)_P \overset{\text{def}}{=} \text{Tr} \ PBA^*,$$ (11)

$$\langle A, B \rangle_P \overset{\text{def}}{=} \text{Re}(A, B)_P = \frac{1}{2}\{(A, B)_P + (B, A)_P\}$$ (12)

$$(A, B \in \mathcal{L}).$$

3. Quantum Cramér-Rao Inequalities

Suppose that a one-parameter model $\{P_\theta \mid \theta \in \Theta\} \subset \mathcal{P}$ of regular statistical operators is given. Then, let us consider the problem of estimating $\theta$ in the situation where the true state lies in the model $\{P_\theta\}$ and $\theta$ is unknown. Since the estimation is given based on a certain measurement, an estimator itself is an observable and denoted by a Hermitian operator. An estimator $T \in \mathcal{L}_h$ is called an unbiased estimator iff it satisfies

$$E_\theta[T] \overset{\text{def}}{=} \text{Tr} \ P_\theta T = \theta \ (\forall \theta \in \Theta).$$ (13)

Then, the variance of $T$ is represented as follows:

$$V_\theta[T] = \text{Tr} \ P_\theta (T - \theta)^2.$$ (14)
Here, for each \( \theta \in \Theta \), from the positivity of \( P_\theta \), there exists unique \( L_\theta = \tilde{L}_\theta \in \mathcal{L}_{s} \) satisfying
\[
\frac{d}{d\theta} P_\theta = \frac{1}{2}(P_\theta L_\theta + L_\theta P_\theta). \tag{15}
\]
This \( L_\theta \) is called the *symmetric logarithmic derivative* of the model \( \{P_\theta\} \) at \( \theta \). In addition,
\[
\tilde{L}_{\theta} \overset{\text{def}}{=} P_{\theta}^{-1} \cdot \frac{d}{d\theta} P_{\theta} \tag{16}
\]
is called the *right logarithmic derivative*. For them, let
\[
I(\theta) \overset{\text{def}}{=} \text{Tr} P_\theta (L_\theta)^2, \tag{17}
\]
\[
\tilde{I}(\theta) \overset{\text{def}}{=} \text{Tr} P_\theta \tilde{L}_\theta \tilde{L}_\theta^*, \tag{18}
\]
then, for any unbiased estimator \( T \),
\[
V_\theta[T] \geq 1/I(\theta), \tag{19}
\]
\[
V_\theta[T] \geq 1/\tilde{I}(\theta) \tag{20}
\]
——Quantum Cramér-Rao inequality—— hold \([1, 2, 3]\). In fact, since
\[
I(\theta) \leq \tilde{I}(\theta) \tag{21}
\]
always holds \([1, 2]\), if we have (19), then we do not need (20). However, in the case when the dimension of the parameter \( \theta \) is more than one, an example is known such that the inequality based on the right logarithmic derivative gives a better evaluation than that based on the symmetric logarithmic derivative \([1, 2, 3]\).

Now, we slightly generalize (19)–(21) and prove them. Let \( \mathcal{K}_\theta \) be the set of \( K_\theta \in \mathcal{L} \) satisfying the following equation:
\[
\frac{d}{d\theta} P_\theta = \frac{1}{2}(P_\theta K_\theta + K_\theta^* P_\theta). \tag{22}
\]
Both \( L_\theta \) and \( \tilde{L}_\theta \) belong to \( \mathcal{K}_\theta \). Hence, it is enough to show that, for any \( K_\theta \in \mathcal{K}_\theta \),
\[
V_\theta[T] \geq 1/\text{Tr}P_\theta K_\theta K_\theta^*, \quad \text{and} \tag{23}
\]
\[
\text{Tr}P_\theta K_\theta K_\theta^* \geq I(\theta). \tag{24}
\]
First, since it holds from (22) that
\[
\langle T - \theta, K_\theta \rangle_{P_\theta} = \text{Tr} \left( \frac{d}{d\theta} P_\theta \right) (T - \theta) = \frac{d}{d\theta} = 1,
\]
Schwarz’s inequality yields that
\[
\langle T - \theta, T - \theta \rangle_{P_\theta} \cdot \langle K_\theta, K_\theta \rangle_{P_\theta} \geq 1.
\]
This is exactly (23). Next, we observe that the equation (22) is equivalent to the following condition:
\[
\langle K_\theta, X \rangle_{P_\theta} = \text{Tr} \left( \frac{d}{d\theta} P_\theta \right) X \text{ for } \forall X \in \mathcal{L}_h.
\]
In the right-hand side of the above equation \( K_\theta \) does not appear. This means that \( K_\theta \) and \( \mathcal{L}_h \) are orthogonal (with respect to the inner product \( \langle \cdot \rangle_{P_\theta} \)). Hence, the orthogonal projection of \( \forall K_\theta \in K_\theta \) to \( \mathcal{L}_h \) is the intersection of \( K_\theta \) and \( \mathcal{L}_h \), that is \( L_\theta \). Hence, we obtain the following:
\[
\langle K_\theta, K_\theta \rangle_{P_\theta} \geq \langle L_\theta, L_\theta \rangle_{P_\theta}.
\]
This is exactly (24).

It follows from the above result that the symmetric logarithmic derivative \( L_\theta \) gives the best lower bound among the logarithmic derivatives \( K_\theta \) satisfying (22). Thus, in what follows, we call \( I(\theta) \), which is defined from \( L(\theta) \) by (17), the Fisher information of the model \( \{P_\theta\} \), and will investigate its properties.

4. Quantum Exponential Family

An unbiased estimator attaining the lower bound in (19) is called an efficient estimator of \( \theta \). We shall clarify the condition under which there exists an efficient estimator. First, from the argument of the previous section, if the equality in (19) holds, then, for each \( \theta \in \Theta \), there exists a constant \( c(\theta) \) satisfying the following:
\[
L_\theta = c(\theta)(T - \theta).
\]
Then, we can integrate (15) as follows:
\[
P_\theta = e^{\frac{1}{2} \{\beta(\theta)T - \gamma(\theta)\}} P_0 e^{\frac{1}{2} \{\beta(\theta)T - \gamma(\theta)\}},
\]
where
\[
\beta(\theta) \overset{\text{def}}{=} \int_0^\theta c(\theta')d\theta', \quad \gamma(\theta) \overset{\text{def}}{=} \int_0^\theta \theta' c(\theta')d\theta'.
\]
Here, $0 \in \Theta$ is assumed for convenience. Conversely, if the model $\{P_\theta\}$ can be written in the form (25), then $T$ can be shown to be an efficient estimator of the parameter $\bar{\theta}$ defined by

$$\bar{\theta} \overset{\text{def}}{=} \frac{d\gamma}{d\beta} = \frac{d\gamma}{d\beta}.$$

Hence, we obtain the following.

**Theorem 1:** For a model $\{P_\theta\}$, an efficient estimator of some parameter exists, if and only if $\{P_\theta\}$ can be written in the form (25).

Equation (25) is an extension of the exponential family in the classical case to the quantum system, and we call it a *quantum exponential family*. In (25), letting $T = \sum_k t_k \pi_k$ be the spectral decomposition of $T$, the distribution $p_\theta(t_k | T) = \text{Tr}P_\theta \pi_k$ of the outcome of the measurement of $T$ can be written as

$$p_\theta(t_k | T) = p_0(t_k | T)e^{\beta(\theta)t_k - \gamma(\theta)},$$

which forms a classical exponential family.

5. Measurements of the Quantum System and Fisher Information

When a measurement $\Pi = \{\pi_k\}$ is fixed, the probability distribution of the outcome of the measurement under a state $P_\theta$ is

$$p_\theta(k | \Pi) \overset{\text{def}}{=} \text{Tr}P_\theta \pi_k. \quad (26)$$

Then, $\{p_\theta(\cdot | \Pi) : \theta \in \Theta\}$ is a classical statistical model, and the accuracy of the parametric estimation can be evaluated by the following Fisher information:

$$I(\theta | \Pi) \overset{\text{def}}{=} \sum_k p_\theta(k | \Pi) \left( \frac{d}{d\theta} p_\theta(k | \Pi) \right)^2. \quad (27)$$

Now, let us study the relation between this classical $I(\theta | \Pi)$ and the quantum $I(\theta)$. First, let

$$\mathcal{M}(\Pi) \overset{\text{def}}{=} \left\{ \sum_k a_k \pi_k \right\} \subset \mathcal{L},$$

$$\mathcal{M}_h(\Pi) \overset{\text{def}}{=} \mathcal{M}(\Pi) \cap \mathcal{L}_h = \left\{ \sum_k a_k \pi_k \right\} \subset \mathcal{L}_h.$$
The set $\mathcal{M} = \mathcal{M}(\Pi)$ satisfies the following.

(i) For $\forall X, \forall Y \in \mathcal{M}$, $\forall a, \forall b \in \mathbb{C}$,

$$aX + bY \in \mathcal{M}, \quad XY \in \mathcal{M}, \quad X^* \in \mathcal{M}.$$ 

(ii) $\text{Id} \in \mathcal{M}$.

In general, an $\mathcal{M}$ satisfying (i) and (ii) above is called a $\ast$-subalgebra of $\mathcal{L}$.

(iii) For $\forall X, \forall Y \in \mathcal{M}$,

$$XY = YX.$$

Moreover, $\mathcal{M}_h(\Pi)$ is the set of observables whose values are determined by the measurement $\Pi$.

Here, we define $L_\theta(\Pi) \in \mathcal{M}_h(\Pi)$ by the following equation:

$$L_\theta(\Pi) \overset{\text{def}}{=} \sum_k \frac{d}{d\theta} \log p_\theta(k | \Pi) \cdot \pi_k.$$ 

Then, for $\forall k$, we have

$$\langle L_\theta, \pi_k \rangle_{P_\theta} = \text{Re} \text{Tr} P_\theta L_\theta \pi_k = \frac{1}{2} \text{Tr} (P_\theta L_\theta + L_\theta P_\theta) \pi_k$$

$$\vdots \quad (15) \quad \Rightarrow \quad \text{Tr} \frac{d}{d\theta} P_\theta \pi_k = p_\theta(k | \Pi) \frac{d}{d\theta} \log p_\theta(k | \Pi) = \langle L_\theta(\Pi), \pi_k \rangle_{P_\theta},$$

so that

$$\langle L_\theta, X \rangle_{P_\theta} = \langle L_\theta(\Pi), X \rangle_{P_\theta} \quad (\forall X \in \mathcal{M}_h(\Pi)).$$

Therefore, $L_\theta(\Pi)$ is the orthogonal projection of $L_\theta$ to $\mathcal{M}_h(\Pi)$. Here, since $I(\theta) = \langle L_\theta, L_\theta \rangle_{P_\theta}$ and $I(\theta | \Pi) = \langle L_\theta(\Pi), L_\theta(\Pi) \rangle_{P_\theta}$, we finally obtain the following.

**Theorem 2:** For any measurement $\Pi$, the following hold for each $\theta \in \Theta$:

$$I(\theta) \geq I(\theta | \Pi), \quad \text{and}$$

$$I(\theta) = I(\theta | \Pi) \iff L_\theta \in \mathcal{M}_h(\Pi).$$

**Corollary 3:** It holds that $I(\theta) = \max_\Pi I(\theta | \Pi)$.

6. Quasi-Classical Model

From the result of the previous section, we see the following. That is, when the true distribution is $P_\theta$, the parametric estimation based on a measurement $\Pi_\theta$ satisfying $L_\theta \in \mathcal{M}_h(\Pi_\theta)$ (the spectral decomposition of $L_\theta$ for instance) is optimal. In general, however, since $\Pi_\theta$ changes depending on
\[ \theta \] there is no measurement that is optimal for all \( \theta \). In this point lies the difficulty peculiar to the estimation problem in the quantum system. If we can take \( \Pi_{\theta} \) independent of \( \theta \), namely, if there exists a measurement \( \Pi \) satisfying

\[ L_{\theta} \in \mathcal{M}_h(\Pi) \text{ for } \forall \theta \in \Theta, \quad (29) \]

then

\[ I(\theta) = I(\theta | \Pi) \text{ for } \forall \theta \in \Theta \quad (30) \]

holds (in this case, \( \Pi \) is said to preserve the Fisher information), and it is sufficient with respect to the estimation of \( \theta \) to perform only the measurement \( \Pi \). In this case, if we carry out, for example, the maximal likelihood estimation for the classical model \( \{p_{\theta}(\cdot | \Pi) | \theta \in \Theta\} \) by using the outcome of the measurement \( \Pi \), then the lower bound of Cramér-Rao (19) is attained asymptotically (as the number of times of the measurement \( \to \infty \)). Thus, when there exists \( \Pi \) satisfying (29) (or (30)) we call \( \{P_{\theta}\} \) a quasi-classical model.

Now, suppose that (29) holds. Then, since it holds that \( L_{\theta}L_{\theta'} = L_{\theta'}L_{\theta} \) \((\forall \theta, \forall \theta' \in \Theta)\), (15) can be integrated as follows:

\[ P_{\theta} = M_{\theta}P_{0}M_{\theta}, \quad (31) \]

where

\[ M_{\theta} \overset{\text{def}}{=} e^{\frac{1}{2} \int_{0}^{\theta} L_{\theta'} d\theta'} \in \mathcal{M}_h(\Pi). \]

Conversely, if \( \{P_{\theta}\} \) is expressed by a certain \( \{M_{\theta}\} \subset \mathcal{M}_h(\Pi) \) as (31), then we have \( L_{\theta} = 2(\frac{d}{d\theta} M_{\theta})M_{\theta}^{-1} \) so that (29) holds. Hence the following hold.

**Theorem 4:** A measurement \( \Pi \) preserves the Fisher information of a model \( \{P_{\theta}\} \) if and only if there exists \( \{M_{\theta}\} \subset \mathcal{M}_h(\Pi) \) satisfying (31).

**Corollary 5:** A model \( \{P_{\theta}\} \) is quasi-classical if and only if it can be written as (31) by some commutative \( \{M_{\theta}\} \subset \mathcal{L}_h \) (i.e. \( M_{\theta}M_{\theta'} = M_{\theta'}M_{\theta} \forall \theta, \forall \theta' \)). In particular, a quantum exponential family (25) is quasi-classical.

7. Sufficiency of Measurements and Fisher Information

For a classical model, a statistic preserves the Fisher information if and only if it is a sufficient statistic. In this section, we study the relation between the condition that a measurement \( \Pi \) preserves the Fisher information and the sufficiency of \( \Pi \) with respect to \( \{P_{\theta}\} \).
First, we shall explain the concept of conditional expectation in the quantum system [4]. Given a statistical operator \( P \in \mathcal{P} \) and a \(*\)-subalgebra \( \mathcal{M} \) of \( \mathcal{L} \), let \( E_P(\cdot \mid \mathcal{M}) \) denote the orthogonal projection from \( \mathcal{L} \) to \( \mathcal{M} \) with respect to the complex inner product \(( \cdot, \cdot)_P\). Namely, for \( \forall A \in \mathcal{L} \), we have \( E_P(A \mid \mathcal{M}) \in \mathcal{M} \), and

\[
(A, X)_P = (E_P(A \mid \mathcal{M}), X)_P \quad (\forall X \in \mathcal{M}). \tag{32}
\]

In the case when \( P \) and \( \mathcal{M} \) satisfy a certain relation\(^\dagger\), one can show that \( E_P(\cdot \mid \mathcal{M}) \) satisfies desirable properties as conditional expectation. Moreover, it is also known that in the other case there does not exist such a desirable conditional expectation. Nevertheless, in this paper, for any \( P \) and \( \mathcal{M} \) we call \( E_P(\cdot \mid \mathcal{M}) \) the conditional expectation with respect to \(( P, \mathcal{M})\).

Next, we shall describe the sufficiency. When \( E_P(\cdot \mid \mathcal{M}) = E_Q(\cdot \mid \mathcal{M}) \) holds for two statistical operators \( P, Q \in \mathcal{P} \) and a \(*\)-subalgebra \( \mathcal{M} \) of \( \mathcal{L} \), we say \( \mathcal{M} \) is sufficient with respect to \( \{ P, Q \} \). (In [4], it is also required that \( P, Q \) and \( \mathcal{M} \) satisfy the “certain relation” mentioned above.) This property is equivalent to

\[
(A, X)_Q = (E_P(A \mid \mathcal{M}), X)_Q \quad (\forall A \in \mathcal{L}, \forall X \in \mathcal{M}). \tag{33}
\]

Here, letting \( N = P^{-1}Q \), it holds that

\[
(A, X)_Q = (A, NX)_P,
\]

and hence (33) can be rewritten as follows:

\[
(A, NX)_P = (E_P(A \mid \mathcal{M}), NX)_P \quad \text{for } \forall A \in \mathcal{L}, \forall X \in \mathcal{M}. \tag{34}
\]

If \( N \in \mathcal{M} \), then the above equation holds from (32). Conversely, if the above equation holds, we can show that \( N \in \mathcal{M} \) by letting \( X = \text{Id} \) especially. Therefore, \( \mathcal{M} \) is sufficient with respect to \( \{ P, Q \} \) if and only if \( P^{-1}Q \in \mathcal{M} \).

When \( \mathcal{M} = \mathcal{M}(\Pi) \), we shall say a measurement \( \Pi \) is sufficient instead of saying \( \mathcal{M}(\Pi) \) is sufficient. Now, if \( \Pi \) is sufficient with respect to \( \{ P, Q \} \), then \( N = P^{-1}Q \in \mathcal{M}(\Pi) \), and hence we can write

\[
N = \sum_k a_k \pi_k \quad (a_k \in \mathbb{C}).
\]

\(^\dagger\)Additional author’s note: Mentioned here is a relation equivalent to the condition that \( \| E_P(A \mid \mathcal{M}) \| \leq \| A \| \) for \( \forall A \in \mathcal{L} \), where \( \| \cdot \| \) denotes the operator norm; see J. Tomiyama, \textit{Proc. Japan Acad.}, 33, 608-612 (1957).
Then, since
\[
\text{Tr}Q\pi_k = \text{Tr}PN\pi_k = a_k\text{Tr}P\pi_k
\]
for each \(k\), we see that \(a_k\) is a positive real number. Hence, the following holds:
\[
N = N^* > 0. \tag{35}
\]
Since \(Q = PN\) is Hermitian, \(PN = NP\) and \(PQ = QP\) hold from (35).

Now, we shall return to the subject of a model \(\{P_\theta\}\). In this case, also, if we say \(\Pi\) is sufficient with respect to \(\{P_\theta\}\) when the conditional expectation \(E_{P_\theta}(\cdot | M(\Pi))\) is independent of \(\theta\), then the following holds.

**Theorem 6:** The condition that a measurement \(\Pi\) is sufficient with respect to \(\{P_\theta\}\) is equivalent to the existence of \(\{N_\theta\} \subset M(\Pi)\) satisfying the following equation:
\[
P_\theta = P_0 N_\theta \quad (\forall \theta \in \Theta). \tag{36}
\]
Moreover, from (36), the following relations are derived:
\[
0 < N_\theta \in M_h(\Pi), \tag{37}
\]
\[
P_\theta N_{\theta'} = N_{\theta'} P_\theta, \tag{38}
\]
\[
P_\theta P_{\theta'} = P_{\theta'} P_\theta \tag{39}
\]
\((\forall \theta, \forall \theta' \in \Theta)\).

In general, a model \(\{P_\theta\}\) is called commutative if it satisfies (39). This condition is equivalent to the existence of a measurement \(\Pi'\) satisfying \(\{P_\theta\} \subset M_h(\Pi')\). Although Theorem 6 does not necessarily implies \(\Pi = \Pi'\), \(\Pi'\) also turns out to be sufficient with respect to \(\{P_\theta\}\). (\(\Pi'\) satisfies the definition of sufficiency in Ref. [4] also.) Hence, the following holds.

**Corollary 7:** There exists a measurement that is sufficient with respect to a model \(\{P_\theta\}\) if and only if the model is commutative.

Now, if \(\Pi\) is sufficient with respect to \(\{P_\theta\}\), then, from (36)-(38), the symmetric logarithmic derivative \(\{L_\theta\}\) of \(\{P_\theta\}\) is given by
\[
L_\theta = \left(\frac{d}{d\theta}N_\theta\right) N_\theta^{-1} \in M_h(\Pi),
\]
so that it satisfies (29). Hence, we obtain the following.

**Theorem 8:** If a measurement \(\Pi\) is sufficient with respect to a model \(\{P_\theta\}\), then \(\Pi\) preserves the Fisher information of \(\{P_\theta\}\).
On the other hand, it is easily seen that the converse of Theorem 8 does not hold. For, a quasi-classical model (31) is commutative only in a special case. For example, in the quantum exponential family (25), the measurement Π of the efficient estimator T preserves the Fisher information, but Π is not sufficient with respect to \( \{ P_\theta \} \) if \( T P_0 \neq P_0 T \). From this, we see that the sufficiency is a quite stronger condition than the preservation of the Fisher information. It is significantly different from the result in the classical case and inditates that we need deeper consideration of the meaning of the condition that the Fisher information is preserved.

8. Conclusion

Throughout the paper, we considered the Fisher information of a quantum statistical model which has one parameter, mainly focusing on its relations with measurement. As future issues, first, by extending the result of this paper, we could discuss the Fisher information, more generally, in the case where a measurement is restricted to an arbitrary \(*\)-subalgebra of \( \mathcal{L} \). In addition, as a more important problem, there remains the extension to the case where the number of parameters is more than one. Since this concerns the problem of simultaneous measurement of non-commuting observables, it is much more difficult than the one-parameter case, and even on the Cramér-Rao inequality, no unified theory has been established so far [1, 2, 3].

References

CHAPTER 10
On the Parameter Estimation Problem for Quantum Statistical Models

Hiroshi Nagaoka

Abstract. The parameter estimation problem for a quantum statistical model $S = \{S_\theta; \theta = (\theta^1, \ldots, \theta^n) \in \Theta\}$, where $\{S_\theta\}$ are quantum states (i.e. density operators) and $\Theta$ is an open set in $\mathbb{R}^n$, is considered. The problem is to estimate the true state which is unknown but is assumed to lie in $S$. The quantum analogues of various notions developed in the classical statistical estimation theory are studied; e.g., unbiased estimator, lower bound of Cramér-Rao type, efficient estimator, asymptotic efficiency, differential geometrical structures, etc. It is shown that the quantum theory for the one-parameter case ($n = 1$) is constructed almost in parallel with the classical theory, while several difficulties lie ahead of us to construct the general theory of quantum estimation for the multi-parameter case.

1. Introduction

Motivated mainly by the recent progress in the field of optical communications, the quantum parameter estimation problem has been actively studied. The problem is to estimate the ‘true’ state which is unknown but is assumed to be represented by a density operator in a quantum statistical model $S = \{S_\theta; \theta = (\theta^1, \ldots, \theta^n) \in \Theta\}$. This can be regarded as a quantum analogue of the parameter estimation problem for a ‘classical’ statistical model $\mathcal{P} = \{p_\theta; \theta = (\theta^1, \ldots, \theta^n) \in \Theta\}$, where $\{p_\theta\}$ are probability density functions, and many notions in the classical theory can find their analogues in the quantum theory. In this paper we study some of these notions such as unbiased estimator, lower bound of Cramér-Rao type, efficient estimator, exponential family, asymptotic efficiency, differential geometrical structures of a model, etc. from a comparative viewpoint between the classical case and the quantum case. It will be shown that the quantum theory for the one-parameter case ($n = 1$) is constructed almost in parallel with the classical theory, while several difficulties lie ahead of us to construct the general quantum theory for the multi-parameter case.

In this paper, we do not delve into the mathematically rigorous treatment of operators on a Hilbert space, but often proceed in an ‘intuitive’
manner, assuming several regularity conditions tacitly. For the rigorous argument, see [6].

8. Efficient Estimator and Exponential Family*

For an $n$-parameter classical statistical model $\mathcal{P} = \{ p_\theta ; \theta \in \Theta \subset \mathbb{R}^n \}$ of probability density functions on $(\Omega, \mathcal{B})$, an estimator $\hat{\theta} : \Omega \to \mathbb{R}^n$ is said to be efficient if $\hat{\theta}$ is unbiased and achieves the Cramér-Rao bound (i.e., its covariance matrix coincides with the inverse of the Fisher information matrix) for $\forall \theta \in \Theta$. It should be noted that the unbiasedness and the efficiency of an estimator depend on the parametrization of $\mathcal{P}$. If $p_\theta$ in $\mathcal{P}$ is written as

$$p_\theta(\omega) = \alpha(\omega) \exp[\sum_{j=1}^n \beta_j(\theta)t_j(\omega) - \gamma(\theta)] \quad (1)$$

by $n + 1$ functions $\{\beta_1, \ldots, \beta_n, \gamma\}$ defined on $\Theta$ and $n + 1$ functions $\{\alpha, t_1, \ldots, t_n\}$ defined on $\Omega$, $\mathcal{P}$ is called an exponential family. In this case, $t = (t_1, \ldots, t_n)$ becomes an efficient estimator for the parameter $\tau = (\tau_1, \ldots, \tau_n)$ defined by

$$\tau_j(\theta) \overset{\text{def}}{=} \int t_j(\omega)p_\theta(\omega)d\omega = \frac{\partial \gamma}{\partial \beta_j}.$$ 

Conversely, if $\hat{\theta}$ is an efficient estimator for a parameter $\theta$ of a model $\mathcal{P}$, then $\mathcal{P}$ is written as (1) where $t_j = \hat{t}_j$ ($j = 1, \ldots, n$). Therefore the necessary and sufficient condition for $\mathcal{P}$ to have an efficient estimator for some parametrization is that $\mathcal{P}$ is an exponential family. These are well-known results in the classical estimation theory.

The corresponding problem in the quantum case is more complicated because we must consider the weight matrix $G$. It should be noted that there is no reason to restrict $G$ to be constant w.r.t. $\theta$. So we suppose that a weight field $\mathcal{G} = \{ G_\theta ; \theta \in \Theta \}$ is given in addition to the model $\mathcal{S} = \{ S_\theta ; \theta \in \Theta \}$. In this situation, an estimator $\hat{M}$ is said to be efficient if $\hat{M}$ is unbiased and satisfies

$$\text{tr} G_\theta V_\theta[\hat{M}] = C^\text{MI}_{\theta}(G_\theta) \quad \text{for} \quad \forall \theta \in \Theta$$

*Editorial note: Section 2 - Section 7 are preliminary parts, where the author gave basic definitions and reviewed the results obtained in his previous paper “A New Approach to Cramér-Rao Bounds for Quantum State Estimation,” which is included in this volume as Chap. 8. Since definitions and notations of this paper are same as those of the previous paper, readers can understand this paper without these sections by referring to Chap. 8 if necessary. Therefore, they are omitted with author’s permission.
where $C^\text{MI}_\theta(G_\theta)$ denotes the most informative CR bound w.r.t. $G_\theta$ at $\theta$.

In the one parameter case $n=1$, however, the efficiency of an estimator does not depend on $G$ as in the classical case. Furthermore, the notion of one parameter exponential family has its quantum analogue as follows. Let $S = \{S_\theta; \theta \in \Theta \subset \mathbb{R}\}$ be a one parameter quantum model. Tracing carefully the derivation of the inequality $V_\theta[\hat{M}] \geq C^S$, we can see that the condition for an estimator $\hat{M}$ to achieve the CR bound $C^S$ for $\forall \theta \in \Theta$, which is equivalent to the efficiency of $\hat{M}$, is that $\hat{M}$ is the measurement of an observable $T$ satisfying

$$
L^S_\theta = c(\theta)(T - \theta), \quad \forall \theta \in \Theta.
$$

(2)

where $L^S_\theta$ is the symmetric logarithmic derivative such that

$$
\frac{\partial}{\partial \theta_j} S_\theta = \frac{1}{2}(S_\theta L^S_\theta + L^S_\theta S_\theta)
$$

(3)

and $c(\theta)$ is a constant. If (2) holds, the differential equation (3) for $S_\theta$ is solved as

$$
S_\theta = \exp\left\{\frac{\beta(\theta)T - \gamma(\theta)}{2}\right\} A \exp\left\{\frac{\beta(\theta)T - \gamma(\theta)}{2}\right\}
$$

(4)

where

$$
\beta(\theta) \stackrel{\text{def}}{=} \int_{\theta_0}^\theta c(\sigma)d\sigma, \quad \gamma(\theta) \stackrel{\text{def}}{=} \int_{\theta_0}^\theta \sigma c(\sigma)d\sigma, \quad A \stackrel{\text{def}}{=} S_{\theta_0}.
$$

Conversely, if $S_\theta$ in $S$ is written as (4) by functions $\{\beta, \gamma\}$ defined on $\Theta$ and Hermitian operators $\{A, T\}$ with $A > 0$, we have

$$
L^S_\theta = \frac{d\beta}{d\theta} T - \frac{d\gamma}{d\theta},
$$

and therefore the measurement of $T$ is an efficient estimator for the parameter

$$
\tau \stackrel{\text{def}}{=} \text{Tr} S_\theta T = \frac{d\gamma}{d\beta}.
$$

A one parameter model $S$ of the form (4) is called a quantum exponential family.

In the multi-parameter case, the condition for a pair $(S, G)$ to have an efficient estimator is not known yet, although there are several important examples of $(S, G)$ for which the existence of efficient estimators have been shown [4, 6].
9. Asymptotic Efficiency

Consider the problem of estimating \( \theta \) under the assumption that observations can be made on a number \( N \) of independent, identical systems, all with the same but unknown state \( S_\theta \) in the model \( \mathcal{S} \). An estimation based on \( N \) observations is represented by a measurement \( \hat{M}^{(N)} \) on \( \mathbb{R}^n \) over the Hilbert space \( \mathcal{H}^{(N)} \overset{\text{def}}{=} \mathcal{H} \otimes \cdots \otimes \mathcal{H} \), the tensor product of \( N \) copies of \( \mathcal{H} \).

In the asymptotic theory of classical statistical estimations, the following results are well-known. Suppose that an estimator \( \hat{\theta}^{(N)} \) based on \( N \) observations for a model \( \mathcal{P} = \{ p_\theta; \theta \in \Theta \} \) is consistent in the sense that, for \( \forall \theta \in \Theta \), \( \hat{\theta}^{(N)} \) converges to \( \theta \) as \( N \to \infty \) in probability w.r.t. \( p_\theta \). Then, under some regularity conditions, the covariance matrix \( V_\theta[\hat{\theta}^{(N)}] \) always satisfies

\[
\lim_{N \to \infty} NV_\theta[\hat{\theta}^{(N)}] \geq J_\theta^{-1}
\]

for \( \forall \theta \in \Theta \), where \( J_\theta \) is the Fisher information matrix of \( \mathcal{P} \) at \( \theta \). This can be regarded as an asymptotic version of the Cramér-Rao inequality, which is based on the fact that the Fisher information matrix for \( N \) observations is given by \( NJ_\theta \). When the equality holds in (5) for \( \forall \theta \in \Theta \), the estimator \( \hat{\theta}^{(N)} \) is said to be (first-order) asymptotically efficient. It can be shown that the maximum likelihood estimator (MLE) is asymptotically efficient for an arbitrary model satisfying some regularity conditions.

Let us proceed to the quantum case. Suppose that a model \( \mathcal{S} = \{ S_\theta; \theta \in \Theta \} \) and a weight field \( \mathcal{G} = \{ G_\theta; \theta \in \Theta \} \) are given, and let \( M_\theta \) be a measurement satisfying \(^1\)

\[
C_\theta^{\text{MI}}(G_\theta) = \text{tr} \ G_\theta (J_\theta M_\theta)^{-1}.
\]

We consider an adaptive estimation procedure as follows: choose the initial value \( \hat{\theta}_0 \in \Theta \) arbitrarily, and then carry out the following (i) and (ii) for \( N = 1, 2, \ldots \) successively. (i) Perform the measurement \( M_{\hat{\theta}_{N-1}} \), and denote the measurement result by \( x_N \). (ii) Using the obtained data \( (x_1, x_2, \ldots, x_N) \), compute the next estimate \( \hat{\theta}_N \) by MLE; i.e., \( \hat{\theta}_N \) is defined as \( \theta' \) maximizing the likelihood function

\[
p_{\theta'}(x_1|\hat{\theta}_0)p_{\theta'}(x_2|\hat{\theta}_1) \cdots p_{\theta'}(x_N|\hat{\theta}_{N-1})
\]

where \( p_{\theta'}(x|\hat{\theta}) \text{d}x = \text{Tr} S_{\theta'} M_{\hat{\theta}} \text{d}x \). It is clear that the procedure defines an estimator \( \hat{M}^{(N)} \) over \( \mathcal{H}^{(N)} \) for \( N = 1, 2, \ldots \), although we do not give

\(^1\)Additional author’s note: See Theorem 2 in Chap. 8 of this volume.
On the Parameter Estimation Problem for Quantum Statistical Models

129

here the explicit expression of $\hat{M}^{(N)}$. When all the quantum system to be observed are in the state $S_\theta$, the probability that $\hat{\theta}_N$ belongs to $B \subset \mathbb{R}^n$ is written as $\text{Tr} S_\theta^{(N)} \hat{M}^{(N)}(B)$, where $S_\theta^{(N)} \overset{\text{def}}{=} S_\theta \otimes \cdots \otimes S_\theta$ (N-tuple).

The consistency of $\hat{M}^{(N)}$, which means that $\hat{\theta}^{(N)}$ tends to $\theta$ as $N \to \infty$, then immediately follows from the consisteny of MLE. Hence, when $N$ is sufficiently large, the computation of $\hat{\theta}^{(N)}$ becomes nearly equal to that of MLE for the model $\{p_\theta(\cdot|\theta); \theta' \in \Theta\}$, whose Fisher information matrices are $\{J_\theta^{M_\theta}; \theta' \in \Theta\}$. Thus, recalling the asymptotic efficiency of MLE and (6), it is expected that the covariance matrix $V_\theta[\hat{M}^{(N)}]$ of the estimator satisfies

$$\lim_{N \to \infty} N \text{tr} G_\theta V_\theta[\hat{M}^{(N)}] = C_{MI}^\theta(G_\theta), \quad \forall \theta \in \Theta. \quad (7)$$

Can this property be called the asymptotic efficiency of the estimator? In the one parameter case, the answer is ‘yes’. Indeed, it is easy to see that, for the model $S^{(N)} \overset{\text{def}}{=} \{S_\theta^{(N)}; \theta \in \Theta\}$, $J_\theta^S$ for $S$ is replaced with $NJ_\theta^S$ [4] and hence $C_{MI}^\theta(G_\theta)$ is replaced with $N^{-1}C_{MI}^\theta(G_\theta)$ (see Section 6†). This proves that the asymptotic performance (7) of $\hat{M}^{(N)}$ is best in the totality of consistent estimators. In the multi-parameter case, however, we do not yet know whether (7) is best or not.

10. Geometrical Structures

A classical statistical model $P = \{p_\theta; \theta \in \Theta\}$ can be regarded as a differential manifold with a coordinate system $\theta$ on which a Riemannian metric $J$ is defined by the Fisher information matrices $\{J_\theta; \theta \in \Theta\}$ and a couple of affine connections, the exponential connection $\Gamma$ and the mixture connection $\tilde{\Gamma}$, are defined in the following way [1]:

$$[\Gamma_\theta]^k_{ij} = \sum_{h=1}^n \int \{\partial_i \partial_j \log p_\theta(\omega)\} \{\partial_h p_\theta(\omega)\} d\omega \cdot [J_\theta^{-1}]^{hk}$$

$$[\tilde{\Gamma}_\theta]^k_{ij} = \sum_{h=1}^n \int \{\partial_i \partial_j p_\theta(\omega)\} \{\partial_h \log p_\theta(\omega)\} d\omega \cdot [J_\theta^{-1}]^{hk}$$

where $\partial_i \overset{\text{def}}{=} \partial / \partial \theta^i$. Using the Kullback information

$$K(p, q) = \int p(\omega) \log \frac{p(\omega)}{q(\omega)} d\omega,$$

†Editorial note: This omitted section explains that $C_{MI} = C^S$ in the one-parameter case; see section 5 of Chap.8 in this book.
these are also written as [2]

\[ [J_{\theta}]_{ij} = -\partial_i \partial_j^\prime K(p_\theta, p_{\theta'})|_{\theta' = \theta}, \quad [\Gamma_{\theta}]^k_{ij} = -\sum_{h=1}^n \partial_i \partial_j^\prime \partial_h K(p_\theta, p_{\theta'})|_{\theta' = \theta} \cdot [J_{\theta}^{-1}]^{hk} \]

\[ [\hat{\Gamma}_{\theta}]^k_{ij} = -\sum_{h=1}^n \partial_i \partial_j \partial_h K(p_\theta, p_{\theta'})|_{\theta' = \theta} \cdot [J_{\theta}^{-1}]^{hk} \]

where \( \partial_i^\prime \overset{\text{def}}{=} \partial / \partial \theta^i \). \( \Gamma \) and \( \hat{\Gamma} \) are mutually dual w.r.t. \( J \) [9, 1] in the sense that

\[ \partial_i [J_{\theta}]_{jk} = \sum_{h=1}^n \{ [\Gamma_{\theta}]^h_{ij} [J_{\theta}]_{hk} + [\hat{\Gamma}_{\theta}]^h_{ik} [J_{\theta}]_{hj} \} . \]

The results in the classical estimation theory mentioned in the previous sections can be understood in this geometrical framework. For instance, an exponential family is characterized as a \( \Gamma \)-autoparallel submanifold of the whole manifold consisting of all the probability density functions which are positive almost everywhere.

A quantum statistical model \( \mathcal{S} = \{ S_\theta ; \theta \in \Theta \} \) can also be regarded as a manifold with a coordinate system \( \theta \). In the sequel we introduce some differential geometrical structures on \( \mathcal{S} \) and examine whether or not they help us to understand the estimation problem for \( \mathcal{S} \).

H. Umegaki [10] defined a quantum analogue of the Kullback information, called the relative entropy, by

\[ K(S, Q) \overset{\text{def}}{=} \text{Tr} S (\log S - \log Q) \]

for density operators \( S \) and \( Q \). Using this \( K \) in the same way as in the classical case, we can introduce on \( \mathcal{S} \) a Riemannian metric \( J^K \) and a couple of affine connections \( \{ \Gamma^K, \hat{\Gamma}^K \} \) which are mutually dual w.r.t. \( J^K \). \( (J^K \) was treated in [7].) However, it seems that the geometrical structure \( (J^K, \Gamma^K, \hat{\Gamma}^K) \) has little connection with the estimation theory. We consider, for instance, a quantum analogue of the notion of exponential family from the viewpoint of this geometry. A one-parameter model \( \mathcal{S} = \{ S_\theta ; \theta \in \Theta \subset \mathbb{R} \} \) is \( \Gamma^K \)-autoparallel (i.e., a \( \Gamma^K \)-geodesic) as a submanifold of the whole manifold consisting of all the nonsingular density operators, if and only if it is written as

\[ S_\theta = \exp[\log A + \beta(\theta)T - \gamma(\theta)] \]

by functions \( \{ \beta, \gamma \} \) defined on \( \Theta \) and Hermitian operators \( \{ A, T \} \) with \( A > 0 \). (As in the classical case, such a model is also characterized as a
set of $S$ minimizing $K(S,Q)$ for some $Q$ under the restriction that $\text{Tr} ST$ has a given value.) It should be noted that (8) is different from the form of quantum exponential family (4) unless $AT = TA$. If $A = I$, both (4) and (8) are reduced to the form of a Gibbs state with the Hamiltonian $T$ appearing in the quantum statistical physics.

The symmetric logarithmic derivatives $\{L^{S}_{\theta,j}; j = 1, \ldots, n\}$ provides us with another geometrical structure $(J^{S}, \Gamma^{S}, \tilde{\Gamma}^{S})$ on $S$ as follows:

$$[J^{S}_{\theta}ij] = \text{Re} \text{Tr} S_{\theta}L^{S}_{\theta,i}L^{S}_{\theta,j} = \text{Tr} (\partial_{i}S_{\theta})(L^{S}_{\theta,j})$$

$$[\Gamma^{S}_{\theta}ik] = \sum_{h=1}^{n} \text{Tr} (\partial_{i}L^{S}_{\theta,j})(\partial_{h}S_{\theta}) \cdot [(J^{S}_{\theta})^{-1}]^{hk}$$

$$[\tilde{\Gamma}^{S}_{\theta}ik] = \sum_{h=1}^{n} \text{Tr} (\partial_{i}\partial_{j}S_{\theta})(L^{S}_{\theta,h}) \cdot [(J^{S}_{\theta})^{-1}]^{hk}.$$  

It is easy to see that $\Gamma^{S}$ and $\tilde{\Gamma}^{S}$ are mutually dual w.r.t. $J^{S}$ and that $\tilde{\Gamma}^{S} = \tilde{\Gamma}^{K}$. We note that $\Gamma^{S}$ is not torsion-free (i.e., $[\Gamma^{S}_{\theta}]^{k}_{ij} \neq [\Gamma^{S}_{\theta}]^{k}_{ji}$) in general and hence, being different from the previously mentioned structures, $(J^{S}, \Gamma^{S}, \tilde{\Gamma}^{S})$ cannot be derived from a nonnegative distance-like function such as $K$. The important fact on this geometry is that a one-parameter quantum exponential family (4) is characterized as a $\Gamma^{S}$-geodesic. Regrettably, the notion of multi-dimensional $\Gamma^{S}$-autoparallel submanifold does not correspond to the existence of an efficient estimator.

Let $\mathcal{G} = \{G_{\theta}; \theta \in \Theta\}$ be a weight field, and denote $J^{G}_{\theta}$ in (6) by $J^{G}_{\theta}$. It can be shown that, if we regard $\mathcal{G}$ as a Riemannian metric on $S$, i.e., if $\mathcal{G}$ is regarded as a covariant tensor field of degree 2 w.r.t. coordinate transformations, then $J^{G} = \{J^{G}_{\theta}; \theta \in \Theta\}$ also becomes a Riemannian metric on $S$. This fact seems to suggest that geometrical considerations will be useful for studying the quantum estimation problem, although at present we do not possess a satisfactory geometrical theory for the problem.

Finally, we note that the following matrix inequality holds in general, whose proof is omitted for want of space.

$$J^{K}_{\theta} \geq J^{S}_{\theta} \geq J^{G}_{\theta} \quad (\forall \theta \in \Theta).$$

References


CHAPTER 11

A Generalization of the Simultaneous Diagonalization of Hermitian Matrices and its Relation to Quantum Estimation Theory

Hiroshi Nagaoka

Abstract. We study the problem of minimizing a quadratic quantity defined for given two Hermitian matrices $X, Y$ and a positive-definite Hermitian matrix. This problem is reduced to the simultaneous diagonalization of $X, Y$ when $XY = YX$. We derive a lower bound for the quantity, and in some special cases solve the problem by showing that the lower bound is achievable. This problem is closely related to a simultaneous measurement of quantum mechanical observables which are not commuting and has an application in the theory of quantum state estimation.

1. Introduction

Suppose that two Hermitian matrices $\{X, Y\} \subset \mathbb{C}^{n \times n}$ and a positive-definite Hermitian matrix $S \in \mathbb{C}^{n \times n}$ are given arbitrarily. For a natural number $m$, we consider $m$ nonnegative-definite Hermitian matrices $\{M_1, \ldots, M_m\} \subset \mathbb{C}^{n \times n}$ and $2m$ real numbers $\{x_1, \ldots, x_m; y_1, \ldots, y_m\} \subset \mathbb{R}$ satisfying

$$\sum_{j=1}^{m} M_j = I \quad \text{(identity)}, \quad (1)$$

$$\sum_{j=1}^{m} x_j M_j = X, \quad \sum_{j=1}^{m} y_j M_j = Y, \quad (2)$$

and define

$$\Delta \overset{\text{def}}{=} \sum_{j=1}^{m} (x_j^2 + y_j^2) \text{Tr}[SM_j] \quad (\geq 0), \quad (3)$$

©1991 Japan Society for Industrial and Applied Mathematics (JSIAM). Reprinted, with permission, from Transactions of Japan Society for Industrial and Applied Mathematics, 1, 43-56, 1991. All or any part of this article may not be reproduced in any form or by any means without permission from JSIAM. This paper was originally written in Japanese. It was translated to English by Fuyuhiko Tanaka and Yoshiyuki Tsuda.
where $\text{Tr}$ denotes the trace. For given $X$, $Y$ and $S$, how can we obtain \{\(M_1, \ldots, M_m\); \(x_1, \ldots, x_m\); \(y_1, \ldots, y_m\)\} minimizing $\Delta$ and the minimum of $\Delta$? Here, no restriction is set to the natural number $m$. That is to say, we wish to minimize $\Delta$ with $m$ allowed to be arbitrarily large (possibly infinite). In this paper, we investigate this problem in relation to the theory of quantum state estimation and solve it in some special cases.

If $X$ and $Y$ are commuting (i.e. $[X,Y] = XY - YX = 0$), then the above problem can be easily solved by the simultaneous diagonalization of $X, Y$ as will be explained later. That is, if we find eigenvalues \{\(x_1, \ldots, x_n\); \(y_1, \ldots, y_n\)\} \(\subset \mathbb{R}\) and eigenvectors \{\(u_1, \ldots, u_n\)\} \(\subset \mathbb{C}^n\) satisfying

\[
Xu_j = x_j u_j, \quad Y u_j = y_j u_j, \quad (4)
\]

\[
u_j^* u_k = \delta_{jk}, \quad (5)
\]

and set

\[
M_j = u_j u_j^* \in \mathbb{C}^{n \times n}, \quad (6)
\]

then \{\(M_j; x_j; y_j\)\} satisfies (1) and (2), and attains $\min \Delta = \text{Tr}[S(X^2 + Y^2)]$, the minimum of $\Delta$.

As it is known well, in the quantum mechanics, an arbitrary physical quantity (which is called an observable) $X$ is represented by a Hermitian operator $X$ on a Hilbert space, and the measurement of $X$ corresponds to the spectral decomposition (namely, diagonalization) of $X$. Moreover, the mathematical fact that two commuting Hermitian operators $X$ and $Y$ can be diagonalized simultaneously is considered as an expression of the physical fact that the corresponding observables $X$ and $Y$ can be measured simultaneously. On the other hand, if $[X,Y] \neq 0$, then $X$ and $Y$ cannot be diagonalized simultaneously, which is considered as a mathematical expression of the uncertainty principle asserting that $X$ and $Y$ cannot be measured simultaneously.

The above fact is written in most of standard textbooks of the quantum mechanics. The concept of measurement there, however, covers only those measurements which can determine the quantum state of the system in arbitrary accuracy, and does not cover all the measurements defined as the operations carried out for the purpose of obtaining a certain information from the physical system. (The former one is sometimes called a simple measurement and the latter a generalized measurement.) Indeed, there are many realistic situations where non-commuting physical quantities such as the position and the momentum are measured simultaneously (sacrificing accuracy to some extent), and they fall into the category of generalized
measurements. Mathematically, the extension from the simple measurement to the generalized measurement corresponds to the one from \( \{ M_j \} \) written by (5), (6) (the set of projections to mutually orthogonal subspaces) to the set of nonnegative-definite Hermitian matrices \( \{ M_j \} \) satisfying only (1) (see Section 3).

Considering such a background, one can easily imagine that the problem mentioned in the first paragraph is related to the simultaneous measurement of non-commuting observables in a certain sense. Indeed, the formulation presented in this paper is a kind of generalization of the simultaneous measurement of two canonical conjugate observables ([1, 3, 5]) which is a well-known concept (see Section 6). However, it is difficult to give a unique definition to the concept of the simultaneous measurement for general non-commuting observables. Instead, we require the formulation where we set a certain criterion and conditions adapted for the purpose of the measurement and find out the optimal measurement under the criterion among those satisfying the conditions. The problem treated in this paper is to find out the measurement minimizing the deviation \( \Delta \) at the given quantum state \( S \) among measurements satisfying the unbiasedness condition (2) on the expectation, and is presented for the purpose of parametric estimation for a quantum statistical model (quantum state estimation) (see Section 3).

The estimation theory for quantum states is a field which has been studied mainly concerned with the optical communications so far, and its formulation can be considered as a kind of extension of the ordinary (classical) statistical estimation problem. Indeed, quantum versions of several concepts of the statistical estimation theory, e.g., Cramér-Rao’s inequality, Fisher information, have been studied and optimal estimation methods have been obtained for some important models as their applications ([7, 3, 5]). However, these results are insufficient in respect of generality, and currently the general theory of quantum estimation corresponding to the classical estimation theory has not been established yet. In constructing such a general theory, the most significant difficulty is that the fundamental mathematical theory on generalized measurements has not been fully developed.

For example, the estimation of complex amplitude of coherent light in background thermal noise ([7]) and its generalization — the estimation of the expectation parameter for the quantum Gaussian state ([5], Chap. VI) — are the most similar cases to the classical estimation theory. However, since the theory is developed deeply depending on the simultaneous measurement theory of canonical conjugate observables, when we try to extend
it to more general case, we face the above-mentioned mathematical problem and get immediately bogged down with it.

The present study is an attempt with the awareness of these issues. If the general theory of quantum estimation is developed, it is expected that, for various situations in which obtaining some information from a quantum system is aimed at, the optimal measurement will be derived mathematically and a wide range of applications will be made possible through the physical consideration on its implementation.

2. General Formulation of Problem

The problem described in Section 1 can be naturally generalized from the following four points of view. (1) To generalize $n \times n$ matrices to operators on a general Hilbert space. (2) To extend from the two $X, Y$ to the case of $X_1, \ldots, X_p$ of arbitrary number. (3) To extend the square sum in the criterion $\Delta$ to a general positive quadratic form. (4) To rewrite the expression of finite $\{M_j; x_j; y_j\}$ to measure-theoretical one including infinite or continuously infinite cases. The most general formulation including the whole is as follows.

Let $\mathcal{H}$ be a Hilbert space and let $X_1, \ldots, X_p$ and $S$ be Hermitian operators on $\mathcal{H}$. Assume that $S$ is positive-definite and its trace is finite. In addition, let $G = [g_{ij}] \in \mathbb{R}^{p \times p}$ be a symmetric matrix of positive-definite. When these are given, we consider a map $M : B \mapsto M(B)$ with the domain $B (\overset{\text{def}}{=} \sigma$-field of all Borel sets of $\mathbb{R}^p)$ and the range of operators on $\mathcal{H}$ satisfying the following conditions.

(i) $\forall B \in \mathcal{B}; \quad M(B) = M(B)^* \geq 0$ (nonnegative-definite)

(ii) $M(\phi) = 0, \quad M(\mathbb{R}^p) = I$ (identity)

(iii) For a sequence of at most countable $\{B_j\} \subset \mathcal{B}$ satisfying $B_i \cap B_j = \phi (\forall i \neq j), \quad M(\bigcup_j B_j) = \sum_j M(B_j)$

(iv) $\int \xi_j M(d\xi) = X_j \quad (d\xi = d\xi_1 \cdots d\xi_p) \quad (j = 1, \ldots, p)$.

If $M(B)$ is not an operator but a scalar, (i)-(iii) is just the definition of a probability measure on $(\mathbb{R}^p, \mathcal{B})$. Therefore, such an $M$ is sometimes called a probability operator measure.* Moreover, the integration of (iv) is defined

*Additional author’s note: This is a synonym for POVM.
just in the same way as the case where an \( M \) is a usual measure. Now, for such an \( M \), our problem is to minimize
\[
\Delta = \Delta(M; S, G) \overset{\text{def}}{=} \sum_{i=1}^{p} \sum_{j=1}^{p} g_{ij} \int \xi_i \xi_j \text{Tr}[SM(d\xi)].
\] (7)

In the above, since \( \mu(d\xi) \overset{\text{def}}{=} \text{Tr}[SM(d\xi)] \) is a measure in the usual sense, \( \Delta \) is always a nonnegative real number. Let \( \Delta^*(X_1, \ldots, X_p; S, G) \) be the minimum of \( \Delta \).

In general, a solution of the problem depends on a weight matrix \( G \). However, if the case of \( G = I_p \) is generally solved, then we can also obtain a solution for arbitrary \( G \). Namely, obtaining \( \Delta^*(X_1, \ldots, X_p; S; G) \) is equivalent to obtaining \( \Delta^*(Y_1, \ldots, Y_p; S; I_p) \), where \( G^{1/2} = [f_{ij}] \) and \( Y_i = \sum_{j=1}^{p} f_{ij} X_j \).

For \( \{M_j; x_j; y_j\} \) satisfying (1) and (2), if we set
\[
M(B) \overset{\text{def}}{=} \sum_{j: (x_j, y_j) \in B} M_j
\]
for any Borel set \( B \subset \mathbb{R}^2 \) and set \( p = 2, X_1 = X, X_2 = Y \), then (i)-(iv) hold. Moreover, in this case, \( \Delta \) in (3) is equal to \( \Delta(M; S; I_2) \). Therefore, the problem in this section is a generalization of the problem in Section 1.

3. Background of Problem

In quantum mechanics, there corresponds a Hilbert space \( \mathcal{H} \) to the quantum system of interest, and various physical concepts for the system are expressed by mathematical concepts for \( \mathcal{H} \). For example, an observable is expressed by a Hermitian operator on \( \mathcal{H} \) and a state of the system is expressed by a nonnegative-definite Hermitian operator of trace one (which is called a density operator or a statistical operator). On the other hand, a (generalized) measurement performed to the system is expressed by a probability operator measure satisfying (i)-(iii) in Section 2. A little more precise description is as follows. A measurement for a quantum mechanical system has the following two aspects. One is the outcome of the measurement, generally considered as a statistic according to a probability distribution. This probability distribution depends on the state \( S \) just before the measurement. The other aspect of a measurement is the state change of the system \( S \rightarrow S' \) caused by the measurement. The mathematical concept describing the former one is the probability operator measure (see, for example, [3, 5, 4]). Now, suppose that a measurement takes values on \( \mathbb{R}^p \). For such
a measurement, there corresponds a probability operator measure \( M \) on \((\mathbb{R}^p, \mathcal{B})\), and the probability distribution of the outcome of the measurement is given by

\[
P(B) = \text{Tr}[SM(B)] \quad (\forall B \in \mathcal{B}),
\]

where \( S \) is a density operator describing the state of the system just before the measurement. It can be easily checked from (i)-(iii) in Section 2 and the fact that \( S \) is a density operator \((S = S^* \geq 0 \text{ and } \text{Tr}[S] = 1)\) that the above \( P \) is a probability measure on \((\mathbb{R}^p, \mathcal{B})\).

A \( p \)-parameter family of density operators \( S = \{S_\theta; \theta = (\theta_1, \ldots, \theta_p) \in \Theta\} \) (\( \Theta \) is an open set of \( \mathbb{R}^p \)) is called a \( p \)-dimensional quantum statistical model. For the family \( S \), a parameter estimation problem can be considered in the same way as one in the ordinary statistical theory. That is, assuming that the true state of the observed physical system belongs to the model \( S \), our problem is to find an estimator which estimates the unknown parameter \( \theta \) as accurately as possible. In this case, an estimator is a measurement with values in \( \mathbb{R}^p \), so that it is described by a probability operator measure \( M \) on \((\mathbb{R}^p, \mathcal{B})\).

When the state of the system is \( S_\theta \), the probability distribution of the estimator can be written as \( P_\theta(B) \stackrel{\text{def}}{=} \text{Tr}[S_\theta M(B)] \quad (\forall B \in \mathcal{B}) \). When the equation

\[
\int \xi_j P_\theta(d\xi) = \theta_j \quad (\forall j \in \{1, \ldots, p\})
\]

holds for \( \forall \theta \in \Theta \), the estimator is called unbiased. With this equation the variance-covariance matrix \( V(\theta) = [v_{ij}(\theta)] \in \mathbb{R}^{p \times p} \) is represented as

\[
v_{ij}(\theta) = \int (\xi_i - \theta_i)(\xi_j - \theta_j) P_\theta(d\xi) = \int \xi_i \xi_j \text{Tr}[S_\theta M(d\xi)] - \theta_i \theta_j.
\]

Now, let us give a real symmetric matrix \( G = [g_{ij}] \in \mathbb{R}^{p \times p} \) as a weight and regard

\[
\text{tr}[GV(\theta)] = \sum_{i=1}^{p} \sum_{j=1}^{p} g_{ij} v_{ij}(\theta) = \Delta(M; S_\theta; G) - \sum_{i=1}^{p} \sum_{j=1}^{p} g_{ij} \theta_i \theta_j
\]

as a quantity representing the deviation of the estimated value (from the true value \( \theta \)) with respect to the estimator \( M \). (The trace \( \text{tr} \) of \( p \times p \) matrices is distinguished from the trace \( \text{Tr} \) of operators on \( \mathcal{H} \).)

Obtaining a lower bound of \( \text{tr}[GV(\theta)] \) for unbiased estimators is equivalent to obtaining a quantum version of Cramér-Rao inequality in the ordinary statistics and plays an important role in the theory of quantum
estimation. (As a matter of fact, we consider another condition called the local unbiasedness at $\theta$, which is a weaker and localized version of the unbiasedness, to develop the theory more simply, although we do not explain the detail here. See [5,6].) For an estimator $M$, when we define Hermitian operators $\{X_1, \ldots, X_p\}$ by Equation (iv) in Section 2, unbiasedness (9) is described by
\[
\text{Tr}[S\theta X_j] = \theta_j \quad (\forall j \in \{1, \ldots, p\}).
\] (10)
Therefore, obtaining a lower bound of $\text{tr}[GV(\theta)]$ is reduced to obtaining a lower bound of $\Delta^*(X_1, \ldots, X_p; S_{\theta}; G)$ for $\{X_1, \ldots, X_p\}$ satisfying (10). Thus we find that the problem mentioned in Section 2 is relevant to the theory of quantum estimation.

4. Fundamental Inequality

We again consider the problem in Section 1 in the following. That is to say, we consider the problem of minimizing $\Delta (\overset{\text{def}}{=} \Delta(M; S))$ in (3) for $M \overset{\text{def}}{=} \{M_j; x_j; y_j\}$ satisfying (1) and (2) when $n \times n$ Hermitian matrices $X$, $Y$ and an $n \times n$ positive-definite Hermitian matrix $S$ are given. (The minimum is denoted by $\Delta^*(X, Y; S)$.) For this problem, the following holds.

**Theorem 1:** Let $\{\lambda_1, \ldots, \lambda_n\}$ be the eigenvalues of the matrix $iS^{1/2}[X, Y]S^{1/2}$ $(i^2 = -1)$. Then, we have
\[
\Delta(M; S) (\geq \Delta^*(X, Y; S)) \geq \text{Tr}[S(X^2 + Y^2)] + \sum_{k=1}^{n} |\lambda_k| \overset{\text{def}}{=} \Gamma(X, Y; S).
\] (11)

**Proof:** For each $j \in \{1, \ldots, m\}$, let $z_j \overset{\text{def}}{=} x_j \pm iy_j$ and $Z \overset{\text{def}}{=} X \pm iY$ (each sign chosen in the corresponding order), then $(z_jI - Z)M_j(z_jI - Z)^*$ is a nonnegative-definite Hermitian matrix. Hence,
\[
\sum_{j=1}^{m} (z_jI - Z)M_j(z_jI - Z)^* = \sum_{j=1}^{m} |z_j|^2 M_j - ZZ^* \quad (\leftarrow (1), (2))
\]
\[
= \sum_{j=1}^{m} (x_j^2 + y_j^2)M_j - (X \pm iY)(X \mp iY)
\]
is also a nonnegative-definite Hermitian matrix, and we obtain the matrix inequality
\[ \sum_{j=1}^{m} (x_j^2 + y_j^2)M_j \geq X^2 + Y^2 \pm i[X, Y]. \]
Since \( S = S^* > 0 \) from the assumption, the above formula results in
\[ \sum_{j=1}^{m} (x_j^2 + y_j^2)S^{\frac{1}{2}}M_jS^{\frac{1}{2}} \geq S^{\frac{1}{2}}(X^2 + Y^2)S^{\frac{1}{2}} \pm iS^{\frac{1}{2}}[X, Y]S^{\frac{1}{2}}. \] (12)
Here, since \( iS^{1/2}[X, Y]S^{1/2} \) is a Hermitian matrix, it has eigenvalues \( \{\lambda_1, \ldots, \lambda_n\} (\subset \mathbb{R}) \) and orthonormal eigenvectors \( \{u_1, \ldots, u_n\} (\subset \mathbb{C}^n) \). Multiplying \( u_k^* \) from the left and \( u_k \) from the right in both sides of Equation (12), we have
\[ \sum_{j=1}^{m} (x_j^2 + y_j^2)u_k^*S^{\frac{1}{2}}M_jS^{\frac{1}{2}}u_k \geq u_k^*S^{\frac{1}{2}}(X^2 + Y^2)S^{\frac{1}{2}}u_k + |\lambda_k|, \] (13)
and obtain (11) by summing up both sides with respect to \( k \).

The above theorem holds in the same form even for a general probability operator measure \( M \) on \( (\mathbb{R}^2; \mathcal{B}) \). Moreover, we may expect it to be naturally extended to the case of operators on an infinite-dimensional Hilbert space (under some regularity conditions).

5. The Commutative Case
If \( [X, Y] = 0 \), then \( \Gamma(X, Y; S) = \text{Tr}[S(X^2 + Y^2)] \) holds. In this case, if we define \( M = \{M_j; x_j, y_j\} \) by (4)-(6), then (1) and (2) obviously hold and
\[ \sum_{j=1}^{m} x_j^2M_j = X^2, \quad \sum_{j=1}^{m} y_j^2M_j = Y^2 \]
hold, so that we have \( \Delta(M; S) = \text{Tr}[S(X^2 + Y^2)] \). Therefore, it gives the minimum of \( \Delta \) and \( \Delta_*(X, Y; S) = \Gamma(X, Y; S) \) holds.

The solution \( M \) in this case depends only on \( X \) and \( Y \), and is independent of \( S \). Moreover, this solution also gives the minimum of \( \Delta(M; S; G) \) with respect to any positive-definite weight matrix \( G \in \mathbb{R}^{2 \times 2} \).

6. The Case of Canonical Commutation Relation
We consider two Hermitian operators \( X \) and \( Y \) satisfying the Heisenberg’s canonical commutation relation \([X, Y] = i\mu I \) (\( \mu \) is a positive constant) such
as the position and the momentum. In this case, even though the Hilbert space \( \mathcal{H} \) is necessarily infinite-dimensional, from \( i S^{1/2}[X, Y] S^{1/2} = -\mu S \) we can write \( \Gamma(X, Y; S) = \text{Tr}[S(X^2 + Y^2)] + \mu \text{Tr}[S] \), and Theorem 1 holds as it is. In the following, we will show the existence of a probability operator measure \( M \) attaining the lower bound of Theorem 1.

For arbitrary two real numbers \( x \) and \( y \), we consider a subspace of \( \mathcal{H} \) consisting of \( \psi \) satisfying

\[
(X + iY)\psi = (x + iy)\psi
\]  

(such a \( \psi \) is called a \textit{minimally uncertain state} or a \textit{coherent state} etc.), and let \( E_{x,y} \) be the orthogonal projection from \( \mathcal{H} \) to this subspace. Note that \((X + iY)E_{x,y} = (x + iy)E_{x,y}\) and its conjugate \(E_{x,y}(X - iY) = (x - iy)E_{x,y}\) hold. Then, it is known that the formula

\[
\frac{1}{2\pi \mu} \int E_{x,y} dx \, dy = I
\]  

holds ([2]). This relation is often referred to as the \textit{completeness} or the \textit{over-completeness} of the coherent states. By multiplying \( X + iY \) from the left in both sides of this formula, we have

\[
\frac{1}{2\pi \mu} \int (x + iy)E_{x,y} dx \, dy = X + iY,
\]  

which can be also written as

\[
\frac{1}{2\pi \mu} \int xE_{x,y} dx \, dy = X, \quad \frac{1}{2\pi \mu} \int yE_{x,y} dx \, dy = Y.
\]  

Moreover, if we multiply \( X - iY \) from the right in both sides of (16), from \([X, Y] = i\mu I\), we have

\[
\frac{1}{2\pi \mu} \int (x^2 + y^2)E_{x,y} dx \, dy = X^2 + Y^2 + \mu I.
\]  

(See, for example, Chap. III in [5] for rigorous treatments of these formulae.) From the completeness (15), a probability operator measure \( M \) on \((\mathbb{R}^2, \mathcal{B})\) is defined by \( M(dx \, dy) \overset{\text{def}}{=} E_{x,y} dx \, dy / 2\pi \mu \), and from (17) and (18),

\[
\Delta(M; S) = \text{Tr}[S(X^2 + Y^2)] + \mu \text{Tr}[S] = \Gamma(X, Y; S)
\]  

holds for any density operator \( S \).

From the above, we see that \( \Delta_*(X, Y; S) = \Gamma(X, Y; S) \) holds also in the case of canonical commutation relation. The solution in this case is independent of \( S \) as well as the one in the last section, whereas, when we consider the minimization of \( \Delta(M; S; G) \) in general, the solution depends on \( G \).
The estimation of complex amplitude of coherent light in background thermal noise ([7]) and the estimation of expectation parameter of quantum Gaussian state as its generalization ([5], Chap. VI) can be analyzed in a similar way to the one in the classical estimation theory, for which the optimal unbiased estimators have been obtained. The mathematical basis of the theory developed there can be found in what we mentioned in the present section. That is, for these models, one can obtain concretely the lower bound of the deviation of the estimated value \( \text{tr}[GV(\theta)] \) defined in Section 3 by applying the above-mentioned relation, and since the estimator \( M \) attaining the lower bound does not depend on a state \( S_\theta \), the uniformly optimal estimator is obtained.

7. 2 × 2 Matrix Case

In this section, we prove the following theorem.

**Theorem 2:** In the case of 2 × 2 matrices,

\[ \Delta_*(X, Y; S) = \Gamma(X, Y; S) \]

always holds.

That is, for arbitrarily given 2 × 2 matrices \( X, Y \) and 2 × 2 positive-definite Hermitian matrix \( S \), we construct \( M = \{M_j; x_j y_j\} \) satisfying (1) and (2) such that \( \Delta(M; S) = \Gamma(X, Y; S) \).

First, we remark the following fact.

**Lemma 3:** If \( X, Y, S \) satisfy \( \Delta_*(X, Y; S) = \Gamma(X, Y; S) \), then, \( \Delta_*(X', Y'; S) = \Gamma(X', Y'; S) \) also holds, where \( X' = X + aI \) and \( Y' = Y + bI \) \( (\forall a \in \mathbb{R}, \forall b \in \mathbb{R}) \).

**Proof:** Suppose that \( M = \{M_j; x_j y_j\} \) satisfy (1), (2) and \( \Delta(M; S) = \Gamma(X, Y; S) \). Then, setting \( x'_j = x_j + a, \ y'_j = y_j + b \) and \( M' = \{M_j; x'_j; y'_j\} \), (1) and (2) hold for \( M', X', Y' \), and we can write

\[
\Delta(M'; S) = \sum_j (x'^2_j + y'^2_j) \text{Tr}[SM_j]
\]

\[= \Delta(M; S) + 2a \text{Tr}[SX] + 2b \text{Tr}[SY] + (a^2 + b^2) \text{Tr}[S].\]

Moreover, since \( [X', Y'] = [X, Y] \),

\[
\Gamma(X', Y'; S) = \Gamma(X, Y; S) - \text{Tr}[S(X^2 + Y^2)] + \text{Tr}[S(X'^2 + Y'^2)]
\]

\[= \Gamma(X, Y; S) + 2a \text{Tr}[SX] + 2b \text{Tr}[SY] + (a^2 + b^2) \text{Tr}[S]\]
holds. Therefore, we have $\Delta(M'; S) = \Gamma(X', Y'; S)$.  

From this lemma, without loss of generality, we can assume  

$$\text{Tr}[SX] = \text{Tr}[SY] = 0.$$  

(19)  

Let us make some additional simplifications. First, since the commutative case has been solved in Section 5, here we assume $[X, Y] \neq 0$. Moreover, we can assume $\text{Tr}[S] = 1$ without loss of generality.  

Now, under these assumptions, we introduce  

$$\mathcal{V} \stackrel{\text{def}}{=} \{V \in \mathbb{C}^{2 \times 2} | V = V^* \text{ and } \text{Tr}[SV] = 0\}$$  

$$\mathcal{W} \stackrel{\text{def}}{=} \{aX + bY | a \in \mathbb{R}, b \in \mathbb{R}\},$$  

for which it turns out that $\mathcal{V}$ is a three-dimensional real linear space and $\mathcal{W}$ is its two-dimensional subspace. Moreover, an inner product on $\mathcal{V}$ is defined as follows:  

$$\langle V_1, V_2 \rangle \stackrel{\text{def}}{=} \text{Re(Tr}[SV_1 V_2]) \quad (V_1, V_2 \in \mathcal{V})$$  

(Re means the real part). Setting  

$$\phi(W) \stackrel{\text{def}}{=} \langle W, X \rangle X + \langle W, Y \rangle Y \in \mathcal{W}$$  

(20) for $\forall W \in \mathcal{W}$, $\phi : W \mapsto \phi(W)$ defines a linear transformation on $\mathcal{W}$.

**Lemma 4:** The map $\phi$ is symmetric and positive-definite with respect to the inner product $\langle \cdot, \cdot \rangle$.  

**Proof:** From  

$$\langle \phi(W_1), W_2 \rangle = \langle W_1, X \rangle \langle X, W_2 \rangle + \langle W_1, Y \rangle \langle Y, W_2 \rangle$$  

$$= \langle W_1, \phi(W_2) \rangle,$$  

$\phi$ is symmetric, and from  

$$\langle \phi(W), W \rangle = \langle W, X \rangle^2 + \langle W, Y \rangle^2,$$  

$\phi$ is positive-definite.  

Therefore, $\phi$ has positive eigenvalues $\{\mu_1, \mu_2\}$ and orthonormal eigenvectors $\{U_1, U_2\} \subset \mathcal{W}$. Namely, we have  

$$\phi(U_j) = \mu_j U_j,$$  

(21)  

$$\langle U_j, U_k \rangle = \delta_{jk}.$$  

(22)
Since \( \{U_1, U_2\} \) forms an orthonormal basis of \( \mathcal{W} \), the matrices \( X \) and \( Y \) can be represented as
\[
X = \langle X, U_1 \rangle U_1 + \langle X, U_2 \rangle U_2, \tag{23}
\]
\[
Y = \langle Y, U_1 \rangle U_1 + \langle Y, U_2 \rangle U_2. \tag{24}
\]
Moreover, the following holds:
\[
\mu_j = \langle \phi(U_j), U_j \rangle = \langle U_j, X \rangle^2 + \langle U_j, Y \rangle^2 \quad (\leftarrow (20)). \tag{25}
\]
On the other hand, since \( U_j \) (\( j = 1, 2 \)) are \( 2 \times 2 \) Hermitian matrices, they have eigenvalues \( \{u_{j1}, u_{j2}\} \subset \mathbb{R} \) and, letting \( \{E_{j1}, E_{j2}\} \) be the matrices of orthogonal projections to the corresponding eigenspaces, we have
\[
U_j = \sum_{k=1}^{2} u_{jk} E_{jk}, \tag{26}
\]
\[
\sum_{k=1}^{2} E_{jk} = I. \tag{27}
\]
Moreover, it holds that
\[
U_j^2 = \sum_{k=1}^{2} u_{jk}^2 E_{jk}. \tag{28}
\]
Now, we set for \( j, k \in \{1, 2\} \)
\[
p_j \overset{\text{def}}{=} \sqrt{\mu_j}/(\sqrt{\mu_1} + \sqrt{\mu_2}), \tag{29}
\]
\[
M_{jk} \overset{\text{def}}{=} p_j E_{jk}, \tag{30}
\]
\[
x_{jk} \overset{\text{def}}{=} p_j^{-1} \langle X, U_j \rangle u_{jk}, \tag{31}
\]
\[
y_{jk} \overset{\text{def}}{=} p_j^{-1} \langle Y, U_j \rangle u_{jk}. \tag{32}
\]
**Lemma 5:** We have
\[
\sum_{j=1}^{2} \sum_{k=1}^{2} M_{jk} = I, \tag{33}
\]
\[
\sum_{j=1}^{2} \sum_{k=1}^{2} x_{jk} M_{jk} = X, \quad \sum_{j=1}^{2} \sum_{k=1}^{2} y_{jk} M_{jk} = Y. \tag{34}
\]
**Proof:** Equation (33) is obvious from (27) and \( p_1 + p_2 = 1 \). Equation (34) is also obvious from (26), (23) and (24).

Remark 6: Equations (26) and (30) indicate that the generalized measurement \( \{ M_{jk} \} \) is a randomized measurement where either one of the two observables \( \{ U_1, U_2 \} \) is chosen with the probability \( \{ p_1, p_2 \} \) respectively and the simple measurement of it is performed.

In what follows, we prove the equation \( \Delta(M; S) = \Gamma(X, Y; S) \) for this \( M = \{ M_{jk}; x_{jk}; y_{jk} \} \).

Lemma 7: It holds that
\[
\Delta(M; S) = \mu_1 + \mu_2 + 2\sqrt{\mu_1 \mu_2} = \langle X, X \rangle + \langle Y, Y \rangle + 2\sqrt{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2}.
\]

Proof: From (30), (31) and (28), we have
\[
\sum_{j=1}^{2} \sum_{k=1}^{2} x_{jk}^2 M_{jk} = \sum_{j=1}^{2} \sum_{k=1}^{2} p_j^{-1} \langle X, U_j \rangle^2 u_{jk}^2 E_{jk} = \sum_{j=1}^{2} p_j^{-1} \langle X, U_j \rangle^2 U_j^2.
\]
Similarly, we have
\[
\sum_{j=1}^{2} \sum_{k=1}^{2} y_{jk}^2 M_{jk} = \sum_{j=1}^{2} p_j^{-1} \langle Y, U_j \rangle^2 U_j^2.
\]
By summing them up, we obtain from (25) and (29)
\[
\sum_{j=1}^{2} \sum_{k=1}^{2} (x_{jk}^2 + y_{jk}^2) M_{jk} = \sum_{j=1}^{2} p_j^{-1} \mu_j U_j^2 = \sum_{j=1}^{2} (\mu_j + \sqrt{\mu_1 \mu_2}) U_j^2.
\]
Multiplying \( S \) from the left and taking trace of this equation, we obtain the first equality in the desired equation. (Note that \( \text{Tr}[SU_j^2] = \langle U_j, U_j \rangle = 1 \).) Moreover, choosing \( \{ X, Y \} \) as a basis of \( \mathcal{W} \), the matrix representation of \( \phi \) is given by
\[
\Phi = \begin{bmatrix}
\langle X, X \rangle & \langle X, Y \rangle \\
\langle Y, X \rangle & \langle Y, Y \rangle
\end{bmatrix},
\]
and hence, \( \mu_1 + \mu_2 \) and \( \mu_1 \mu_2 \) are expressed by the trace and the determinant of \( \Phi \) respectively, which proves the second equality.

Lemma 8: Letting \( \{ \lambda_1, \lambda_2 \} \) be the eigenvalues of \( \Lambda \overset{\text{def}}{=} iS^{1/2}[X, Y]S^{1/2} \), we have
\[
\sum_{j=1}^{2} |\lambda_j| = 2\sqrt{\langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2}.
\]
Proof: All the quantities in the above equation are invariant under the transformations $X \rightarrow UXU^*$, $Y \rightarrow UYU^*$ and $S \rightarrow USU^*$ for an arbitrary $2 \times 2$ unitary matrix $U$. Therefore, without loss of generality, we can assume that $S$ is a diagonal matrix of the form

$$S = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}, \quad (s_j > 0, \ s_1 + s_2 = 1).$$

When we define

$$V_1 \overset{\text{def}}{=} \begin{bmatrix} \sqrt{s_2/s_1} & 0 \\ 0 & -\sqrt{s_1/s_2} \end{bmatrix}, \quad V_2 \overset{\text{def}}{=} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad V_3 \overset{\text{def}}{=} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix},$$

they form a basis of $\mathcal{V}$. Therefore, $X (\in \mathcal{V})$ and $Y (\in \mathcal{V})$ can be written as

$$X = \sum_{j=1}^{3} a_j V_j, \quad Y = \sum_{j=1}^{3} b_j V_j.$$

From straightforward computation, the $(j,k)$ component $q_{jk}$ of $Q \overset{\text{def}}{=} S^{1/2}XYS^{1/2}$ is given by

$$q_{11} = s_2 a_1 b_1 + s_1 (a_2 b_2 + a_3 b_3) - is_1 (a_2 b_3 - a_3 b_2),$$

$$q_{12} = s_2 a_1 b_2 - s_1 a_2 b_1 + i(s_2 a_1 b_3 - s_1 a_3 b_1),$$

$$q_{21} = s_2 a_2 b_1 - s_1 a_1 b_2 + i(s_1 a_1 b_3 - s_2 a_3 b_1),$$

$$q_{22} = s_1 a_1 b_1 + s_2 (a_2 b_2 + a_3 b_3) + is_2 (a_2 b_3 - a_3 b_2).$$

Hence, we have

$$\langle X, Y \rangle = \text{Re}(\text{Tr}[Q]) = \sum_{j=1}^{3} a_j b_j = \vec{a} \cdot \vec{b}, \quad (35)$$

where $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$. The above formula holds as long as both $X$ and $Y$ are elements of $\mathcal{V}$, which implies that $\{V_1, V_2, V_3\}$ forms an orthonormal basis of $\mathcal{V}$. Therefore,

$$\langle X, X \rangle = \sum_{j=1}^{3} a_j^2 = \| \vec{a} \|^2, \quad \langle Y, Y \rangle = \sum_{j=1}^{3} b_j^2 = \| \vec{b} \|^2 \quad (36)$$

also hold.

On the other hand, $\Lambda$ can be written as

$$\Lambda = i(Q - Q^*) = \begin{bmatrix} 2s_1 c_1 & c_2 + ic_3 \\ c_2 - ic_3 & -2s_2 c_1 \end{bmatrix},$$
where
\[ c_1 \overset{\text{def}}{=} a_2 b_3 - a_3 b_2, \quad c_2 \overset{\text{def}}{=} a_3 b_1 - a_1 b_3, \quad c_3 \overset{\text{def}}{=} a_1 b_2 - a_2 b_1. \] (37)
The vector \( \vec{c} = (c_1, c_2, c_3) \) equals \( \vec{a} \times \vec{b} \), i.e., the vector product of \( \vec{a} \) and \( \vec{b} \). Since
\[ \lambda_1 \lambda_2 = \det \Lambda = -4 s_1 s_2 c_1^2 - c_2^2 - c_3^2 \leq 0, \]
we have
\[ \sum_{j=1}^{2} |\lambda_j| = |\lambda_1 - \lambda_2| = \sqrt{(\lambda_1 + \lambda_2)^2 - 4\lambda_1 \lambda_2} \] (38)
\[ = \sqrt{(\text{Tr}[\Lambda])^2 - 4 \det \Lambda} = 2 \sqrt{c_1^2 + c_2^2 + c_3^2} = 2 \| \vec{a} \times \vec{b} \|. \]
Letting \( \theta \) be the angle between \( \vec{a} \) and \( \vec{b} \), we can write
\[ \| \vec{a} \times \vec{b} \|^2 = \| \vec{a} \|^2 \cdot \| \vec{b} \|^2 \sin^2 \theta = \| \vec{a} \|^2 \cdot \| \vec{b} \|^2 \left(1 - \cos^2 \theta\right) \]
\[ = \| \vec{a} \|^2 \cdot \| \vec{b} \|^2 - (\vec{a} \cdot \vec{b})^2, \]
so that from (35) and (36) we have
\[ \| \vec{a} \times \vec{b} \|^2 = \langle X, X \rangle \langle Y, Y \rangle - \langle X, Y \rangle^2. \] (39)
From (38) and (39), we obtain the desired equation.

From Lemmas 7 and 8, \( \Delta(M; S) = \Gamma(X, Y; S) \) holds and therefore Theorem 2 has been proved. The above solution generally depends on \( S \), which is different from the cases treated in Sections 5 and 6.

8. Conclusion
For the problem of minimizing \( \Delta \) posed in Section 1, we have obtained \( \Gamma(X, Y; S) \) as a lower bound and showed that the bound is attainable for some cases. From these results, one may conjecture that \( \Delta_*(X, Y; S) = \Gamma(X, Y; S) \) always holds in the case of general \( n \times n \) matrices, or of operators on an arbitrary Hilbert space. So far, the problem of whether this conjecture is true or not has not been solved.

When the method used in Section 7 is applied to the general case of \( n \times n \) matrices, Lemmas 3, 4, 5 and 7 still hold by the same arguments, but Lemma 8 does not hold in general. Therefore, even if we construct \( M \) by the same method, the equation \( \Delta(M; S) = \Gamma(X, Y; S) \) does not necessarily hold. In order to solve the case of \( n \geq 3 \), it seems necessary to essentially extend the argument of Section 7. Indeed, it is very difficult at the current...
stage to find out a meaningful relation between the solution in Section 6 and the one in Section 7. However, if we succeed in constructing a general theory which can be applied to any $n$, the gap will be filled and a new light will be shed on the method used in Section 6. Even for $n \times n$ matrices to which the method of Section 7 cannot be applied, there are some special cases in which it is possible to construct $M$ satisfying $\Delta(M; S) = \Gamma(X, Y; S)$ by some combinatoric trial and error$^\dagger$. Thus, the author would expect that a key to the general theory will be obtained by accumulating such examples of construction.

Though the above arguments are all for two $X$ and $Y$, it is also important to extend them to the case of an arbitrary number of $\{X_1, \ldots, X_p\}$ mentioned in Section 2. For that purpose, we have to first extend Theorem 1 of Section 4. The extension, however, will require a completely new insight since the proof of Theorem 1 is deeply dependent on the condition $p = 2$.

In this paper, we took up a mathematical problem as one of the difficulties arising in constructing the general theory of quantum state estimation and focused only on it. Our argument reveals a mathematical difficulty in generalizing known results for special models. At the same time, it is also an attempt to explore the possibility of overcoming such a difficulty. Of course, the problem taken up here is a mere part of the whole theory of quantum estimation and we can hardly expect meaningful applications only with it. For example, if the problem of this paper is completely solved, then $\Delta_*(X_1, \ldots, X_p; S_\theta; G)$ in Section 3 can be obtained, but to obtain the lower bound of $\text{tr}[GV(\theta)]$ from this, i.e., to derive the Cramér-Rao’s inequality, we further need to obtain the lower bound of $\Delta_*(X_1, \ldots, X_p; S_\theta; G)$ for $\{X_1, \ldots, X_p\}$ satisfying the (local) unbiasedness. (In the case of $2 \times 2$ matrices, the lower bound can be obtained and the author will discuss it at another opportunity$^\ddagger$.) Moreover, since the Cramér-Rao’s inequality gives the lower bound for the deviation with the unknown parameter $\theta$ fixed, it is necessary to clarify in what case the lower bound is uniformly attained independent of the value of $\theta$, namely, what corresponds to the concepts of efficient estimator and of exponential family in the classical estimation theory. Thus, the contribution of this paper to the theory of quantum estimation is not enough by itself, or rather, we may say that its significance is to give a basis

$^\dagger$Dr. Naoki Suehiro showed interesting constructions for some numerical examples of $3 \times 3$ in a private communication to the author.

$^\ddagger$Additional author’s note: See the note at the end of chap. 8 in this book.
to consider the problems mentioned above as well as problems concerning such notions as the asymptotic efficiency and the sufficiency.

References

CHAPTER 12
A Linear Programming Approach to Attainable Cramér-Rao Type Bounds

Masahito Hayashi

Abstract. The author studies the relation between the attainable Cramér-Rao type bound and the duality theorem in the infinite dimensional linear programming. By this approach, the attainable Cramér-Rao type bound for a 3-parameter spin-1/2 model is explicitly derived.

1. Introduction

It is well-known that the lower bound of quantum Cramér-Rao inequality \( V_\rho(M) \geq J^S_\rho \) cannot be attained unless all the SLDs commute, where we denote by \( V_\rho(M) \) a covariance matrix for a state \( \rho \) by a measurement \( M \), the SLD Fisher information matrix for a state \( \rho \) by \( J^S_\rho \). We therefore often treat an optimization problem for \( \text{tr} \ GV_\rho(M) \) to be minimized, where \( G \) is an arbitrary real positive symmetric matrix. If there is a function \( C_\rho \) (possibly depending on \( G \)) such that \( \text{tr} \ GV_\rho \geq C_\rho \) holds for all \( M \), \( C_\rho \) is called a Cramér-Rao type bound, or simply a CR bound. Our purpose is to find the most informative (i.e., attainable) CR bound under locally unbiasedness conditions. To author’s knowledge, there has been known only two models for which this optimization problem was explicitly solved in mixed state models. One is the estimation of complex amplitudes of coherent signals in Gaussian noise solved by Yuen and Lax [8], and Holevo [5]. Another one is the estimation of a 2-parameter spin-1/2 model solved by Nagaoka [6]. In pure state case, see Fujiwara and Nagaoka [1].

In this paper, a completely different approach to the optimization problem is given based on an infinite dimensional linear programming technique [7]. In addition, the most informative CR bound for a 3-parameter spin-1/2 model is explicitly derived by this approach.

2. SLD Inner Product and Locally Unbiasedness Conditions

**Definition 1:** For a subset $\Theta \subset \mathbb{R}^n$ the map $f: \Theta \to T_{sa}(\mathcal{H})$ is called a $C^k$-map, if the $k$-th derivative of $f$ is well defined on the interior of $\Theta$, where $T_{sa}(\mathcal{H})$ is the set of selfadjoint trace class operators on $\mathcal{H}$.

By $T_{sa}^{+\cdot1}(\mathcal{H})$ we denote the set $\{\rho \in T_{sa}(\mathcal{H})|\rho \geq 0, \ tr\mathcal{H}\rho = 1\}$.

**Definition 2:** We call $P \subset T_{sa}^{+\cdot1}(\mathcal{H})$ an $n$-dimensional $C^k$-state manifold, if $P$ is an $n$-dimensional $C^k$-manifold which satisfies the condition: there exists a family $\{U_\lambda\}_{\lambda \in \Lambda}$ of open sets of $P$ such that for any $\lambda \in \Lambda$ there exist $\Theta_\lambda \subset \mathbb{R}^n$ and $C^k$ map $\phi_\lambda : \Theta_\lambda \to U_\lambda$.

We assume that we are given a family of density operators $P$ which is a state manifold. The space $L^2_{sa}(\rho)$ is defined as follows.

**Definition 3:** For $\rho \in T_{sa}(\mathcal{H})$, $L^2_{sa}(\rho)$ consists of selfadjoint operators $X$ on $\mathcal{H}$ satisfying the following conditions:

1) $\phi_j \in \mathcal{D}(X)$ with respect to $j$ such that $s_j \neq 0$  \hspace{1cm} (1)

2) $\langle X|X\rangle^sa_\rho := \sum_j s_j \langle X\phi_j|X\phi_j\rangle < \infty$,  \hspace{1cm} (2)

where $\rho = \sum_j s_j |\phi_j\rangle\langle \phi_j|$ is the spectral decomposition of $\rho$.

**Definition 4:** We call a state $\rho$ faithful, if it satisfies the following condition:

$X \in L^2_{sa}(\rho)$ and $\langle X|X\rangle^sa_\rho = 0 \implies X = 0$  \hspace{1cm} (3)

We assume that all of the given family of density operators are faithful. For $X, Y \in L^2_{sa}(\rho)$, define

$\langle X|Y\rangle^sa_\rho := \frac{1}{4} \left( \langle X+Y|X+Y\rangle^sa_\rho - \langle X-Y|X-Y\rangle^sa_\rho \right)$.  \hspace{1cm} (4)

We denote the norm with respect to this inner product by $\|\|^{S}_{\rho}$. We call the inner product the SLD inner product.

**Theorem 5:** If $\mathcal{H}$ is separable, $L^2_{sa}(\rho)$ is a real Hilbert space with respect to the SLD inner product.

For a proof see Holevo [4].
We define \( \rho \circ X := \frac{1}{2}(\rho \cdot X + X \cdot \rho) \in T_{sa}(\mathcal{H}) \) for \( X \in \mathcal{L}_{sa}^{2}(\rho) \). We denote the inner product of the real Hilbert space \( \mathcal{L}_{sa}^{2}(\rho) \) by \( J_{\rho}^{S} \). We identify the dual of \( \mathcal{L}_{sa}^{2}(\rho) \) with \( \mathcal{L}_{sa}^{2,*}(\rho) := \{ \rho \circ X | X \in \mathcal{L}_{sa}^{2}(\rho) \} \) as follows.

\[
\mathcal{L}_{sa}^{2,*}(\rho) \times \mathcal{L}_{sa}^{2}(\rho) \rightarrow \mathbb{R} \\
(x, X) \mapsto \operatorname{tr}_{\mathcal{H}} x X.
\]

We may regard \( J_{\rho}^{S} \) as an element of \( \text{Hom}_{sa}(\mathcal{L}_{sa}^{2}(\rho), \mathcal{L}_{sa}^{2,*}(\rho)) \) by

\[
J_{\rho}^{S} : \mathcal{L}_{sa}^{2}(\rho) \rightarrow \mathcal{L}_{sa}^{2,*}(\rho) \\
X \mapsto \rho \circ X.
\]

We identify \( \frac{\partial}{\partial \theta} \in T_{P}P \) with \( \frac{\partial}{\partial \theta} \in T_{sa}(\mathcal{H}) \), and we assume that \( T_{P}P \subset \mathcal{L}_{sa}^{2,*}(\rho) \). We denote \( J_{\rho}^{S}|_{T_{P}P} \) by \( J_{\rho}^{P} \). We identify \( T_{P}P \) with \( J_{\rho}^{P,-1}(T_{P}P) \). We call the inner product \( J_{\rho}^{P,-1} \) on \( T_{P}P \) the *SLD inner product. By \( \| \|_{S} \) we denote the norm of *SLD inner product \( J_{\rho}^{P,-1} \). By \( \mathcal{M}(\Omega, \mathcal{H}) \) we denote the set of generalized measurements on \( \mathcal{H} \) whose measurable space is \( \Omega \).

**Definition 6:** We define an affine map \( E \) from \( \mathcal{M}(T_{P}P, \mathcal{H}) \) to \( \text{Hom}(T_{sa}(\mathcal{H}), T_{P}P) \) by

\[
E(M)(\tau) := \int_{T_{P}P} x \operatorname{tr}_{\mathcal{H}} \left( M(dx) \tau \right), \quad \forall \tau \in T_{sa}(\mathcal{H}).
\]

Let us define the locally unbiasedness condition.

**Definition 7:** We call \( M \in \mathcal{M}(T_{P}P, \mathcal{H}) \) locally unbiased on \( \rho \in P \), if the map \( E(M) : T_{sa}(\mathcal{H}) \rightarrow T_{P}P \) satisfies the following condition:

\[
E(M)(\rho) = 0 \\
E(M)|_{T_{P}P} = \text{Id}_{T_{P}P}.
\]

We denote the set of locally unbiased measurements on \( \rho \in P \) by \( \mathcal{U}(T_{P}P) \).

**Lemma 8:** For \( M \in \mathcal{M}(T_{P}P, \mathcal{H}) \), the condition (9) is equivalent to the following equation:

\[
\int_{T_{P}P} \operatorname{tr}_{\mathcal{H}} a(x) M(dx) = \operatorname{tr}_{T_{P}P} a, \quad \forall a \in \text{End}(T_{P}P).
\]

By taking basis, it is easy to verify this.

Let \( g \) be a nonnegative inner product on \( T_{P}P \), then we call \( \inf_{M \in \mathcal{U}(T_{P}P)} \operatorname{tr}_{T_{P}P} V_{P}(M)g \) the attainable Cramér-Rao bound.
3. Linear Programming Approach

We introduce a new approach to the attainable Cramér-Rao type bound. In this approach, we apply the duality theorem of the infinite dimensional linear programming [7]. But, we don’t have to know the duality theorem for this section. In the noncommutative case, there is no infimum of covariance matrices under the locally unbiasedness conditions. Therefore, we minimize the following value $D_\rho^g$ under the locally unbiasedness conditions. Let $g$ be a nonnegative inner product on $T_\rho P$.

**Definition 9:** We define the deviation $D_\rho^g$ for a measurement $M \in \mathcal{M}(T_\rho P, \mathcal{H})$ as follow:

$$D_\rho^g(M) := \text{tr}_{T_\rho P} gV_\rho(M) = \int_{T_\rho P} g(x,x) \text{tr}_\mathcal{H} M(dx) \rho.$$ 

We introduce the useful theorem to minimize the deviation $D_\rho^g(M)$ under the locally unbiasedness conditions.

**Theorem 10:** We have the inequality

$$\inf_{M \in \mathcal{U}(T_\rho P)} D_\rho^g(M) \geq \sup_{(a,S) \in \mathcal{U}^*(g)} (\text{tr}_{T_\rho P} a + \text{tr}_\mathcal{H} S),$$

where

$$\mathcal{U}^*(g) := \{(a,S) \in \text{End}(T_\rho P) \times T_{sa}(\mathcal{H}) | g(x,x) \cdot \rho - S - a(x) \in T^+_sa(\mathcal{H}), \forall x \in T_\rho P\}.$$ 

Notice that $T_\rho P$ is a subset $T_{sa}(\mathcal{H})$.

We call the calculation of $\sup_{(a,S) \in \mathcal{U}^*(g)} (\text{tr}_{T_\rho P} a + \text{tr}_\mathcal{H} S)$ the dual problem.

**Corollary 11:** If there exist a sequence of locally unbiased measurements $\{M_k\}$ and an element $(a',S')$ of $\mathcal{U}^*(g)$ satisfying the condition

$$R_\rho^g(a',S';M_k) \to 0 \text{ (as } k \to 0),$$

then

$$\lim_{k \to \infty} D_\rho^g(M_k) = \text{tr}_{T_\rho P} a' + \text{tr}_\mathcal{H} S' = \inf_{M \in \mathcal{U}(T_\rho P)} D_\rho^g(M)$$

$$= \sup_{(a,S) \in \mathcal{U}^*(g)} (\text{tr}_{T_\rho P} a + \text{tr}_\mathcal{H} S),$$

where $\mathcal{R}_\rho^g$ is defined as

$$\mathcal{R}_\rho^g(a,S;M) := \text{tr}_\mathcal{H} \int_{T_\rho P} R_\rho^g(a,S;x)M(dx)$$

$$R_\rho^g(a,S;x) := g(x,x) \cdot \rho - S - a(x).$$
We call \((a, S) \in U^*(g)\) the Lagrange multiplier.

**Proof of Theorem 10 and Corollary 11:** For \(M \in U(T_\rho P)\) and \((a, S) \in U^*(g)\), we have

\[
R^\rho_g(a, S; M) = \text{tr}_\mathcal{H} \int_{T_\rho P} g(x, x) \cdot \rho M(dx) - \text{tr}_\mathcal{H} \int_{T_\rho P} SM(dx) - \text{tr}_\mathcal{H} \int_{T_\rho P} a(x)M(dx)
\]

\[
= D^\rho_g(M) - \text{tr}_{T_\rho P} a - \text{tr}_\mathcal{H} S. \tag{16}
\]

Since we have \(R^\rho_g(a, S; x) \geq 0\) for any \(x \in T_\rho P\), we obtain \(R^\rho_g(a, S; M) \geq 0\). By (16), the proof is complete.

When \(\dim \mathcal{H} < \infty\), we have the equality in (11). See Hayashi [2,3].

4. Maximum

In this section, we consider the dual problem. Let us define a linear functional on \(\text{End}(T_\rho P) \times T_{sa}(\mathcal{H})\), denoted by \(\text{Tr}\), in the following way, where in this section, we regard \(T_\rho P\) as a real Hilbert space with respect to \(J_{\rho, -1} S, T\).

**Definition 12:** The functional \(\text{Tr}\) is defined by

\[
\text{Tr} : \text{End}(T_\rho P) \times T_{sa}(\mathcal{H}) \to \mathbb{R}
\]

\[
(a, S) \mapsto \text{tr}_{T_\rho P} a + \text{tr}_\mathcal{H} S.
\]

**Lemma 13:** If the dimension of \(\mathcal{H}\) is finite, then the set \(U^*(g) \cap \text{Tr}^{-1}([0, \infty))\) is compact.

We assume that the norm of \(\text{End}(T_\rho P)\) is the operator norm \(\| \|_o\), and the norm of \(T_{sa}(\mathcal{H})\) is the trace norm \(\| \|_t\). We define the norm \(\| \|_{o,t}\) of \(\text{End}(T_\rho P) \times T_{sa}(\mathcal{H})\) as follows:

\[
\|(a, S)\|_{o,t} := \|a\|_o + \|S\|_t, \forall (a, S) \in \text{End}(T_\rho P) \times T_{sa}(\mathcal{H}).
\]

**Proof:** We have

\[
U^*(g) = \cap_{x \in T_\rho P} \{(a, S) | g(x, x) \cdot \rho - S - a(x) \in T_{sa}^+(\mathcal{H})\}.
\]

Moreover, \(\{(a, S) | g(x, x) \cdot \rho - S - a(x) \in T_{sa}^+(\mathcal{H})\}\) is closed. Thus, \(U^*(g)\) is closed. Because \(\text{Tr}^{-1}([0, \infty))\) is closed, \(U^*(g) \cap \text{Tr}^{-1}([0, \infty))\) is closed. Therefore, it is sufficient to show that it is bounded with respect to the norm \(\| \|_{o,t}\). Denote \(n := \dim T_\rho P\). For \((a, S) \in U^*(g) \cap \text{Tr}^{-1}([0, \infty))\), we have
\[ \text{tr}_{\rho} P a \leq n\|a\|_o. \]

Choose \( z \in T_{\rho} P \) such that \( \|z\|_S = 1, \|a(z)\|_S = \|a\|_o. \) For \( r > 0, \) we have

\[
g(r \cdot z, r \cdot z) \rho - a(r \cdot z) - S \geq 0. \tag{17}
\]

Substitute \( r = 0, \) then \(-S \geq 0. \) Let us calculate the left hand side of (17).

\[
g(r \cdot z, r \cdot z) \rho - a(r \cdot z) - S = r^2 \cdot g(z, z) \rho - r \cdot J_{S,T}^{\rho,-1}(a(z)) \circ \rho - S
\]

\[
= \left( \sqrt{g(z, z)} r - \frac{1}{2 \sqrt{g(z, z)}} J_{S,T}^{\rho,-1}(a(z)) \right) \cdot \rho
\]

\[
\cdot \left( \sqrt{g(z, z)} r - \frac{1}{2 \sqrt{g(z, z)}} J_{S,T}^{\rho,-1}(a(z)) \right)
\]

\[
- \frac{1}{4 g(z, z)} J_{S,T}^{\rho,-1}(a(z)) \cdot \rho \cdot J_{S,T}^{\rho,-1}(a(z)) - S. \tag{18}
\]

Let \( \{e_i\} \) be a complete orthonormal system of \( \mathcal{H} \) which consists of eigenvectors of

\[ J_{S,T}^{\rho,-1}(a(z)). \]

Substitute \( r \) for the eigenvalue \( \alpha_i \) of \( \frac{1}{2 g(z, z)} J_{S,T}^{\rho,-1}(a(z)) \) corresponding to the eigenvector \( e_i, \) then we have

\[
\left\langle e_i \left| \sqrt{g(z, z)} \alpha_i - \frac{1}{2 \sqrt{g(z, z)}} J_{S,T}^{\rho,-1}(a(z)) \right. \right. \cdot \rho \cdot \left( \sqrt{g(z, z)} \alpha_i - \frac{1}{2 \sqrt{g(z, z)}} J_{S,T}^{\rho,-1}(a(z)) \right) \left| e_i \right\rangle = 0.
\]

By (17), we have

\[
\left\langle e_i \left| - \frac{1}{4 g(z, z)} J_{S,T}^{\rho,-1}(a(z)) \cdot \rho \cdot J_{S,T}^{\rho,-1}(a(z)) - S \right| e_i \right\rangle \geq 0.
\]

Therefore, we have

\[
\text{tr}_{\mathcal{H}} \left( - \frac{1}{4 g(z, z)} J_{S,T}^{\rho,-1}(a(z)) \cdot \rho \cdot J_{S,T}^{\rho,-1}(a(z)) - S \right) \geq 0.
\]

Thus, we get

\[
\text{tr}_{\mathcal{H}} S \leq - \frac{1}{4 g(z, z)} \langle a(z) | a(z) \rangle^o_S = - \frac{\|a\|_o^2}{4 g(z, z)}, \quad \text{tr}_{\mathcal{H}} S \leq - \frac{\|a\|_o^2}{4 \|g\|_o}.
\]

Therefore, we obtain

\[
0 \leq \text{Tr}(a, S) \leq n\|a\|_o - \frac{\|a\|_o^2}{4 \|g\|_o}.
\]

Hence, \( 0 \leq \|a\|_o(n - \frac{\|a\|_o}{4 \|g\|_o}). \) Thus, \( 0 \leq \|a\|_o \leq 4n\|g\|_o. \) As \(-S \geq 0, \) we have \( \|S\| = -\text{tr}_{\mathcal{H}} S. \) Therefore, we obtain the following inequalities:

\[
0 \leq \|S\| \leq \text{tr} a \leq n\|a\|_o \leq 4\|g\|_o n^2.
\]
Moreover the equality holds, if and only if there exists \( \rho \). We have \( \rho \).

Proof: \( \rho \).

Corollary 14: There exists the maximum of the right hand side of (11).

Assume that \( \rho \in P_1 \subset P_2 \) and \( T_\rho P_1 \subset T_\rho P_2 \), \( T_\rho P_1 \neq T_\rho P_2 \). From the imbedding map \( i : P_1 \hookrightarrow P_2 \), we have \( di_\rho : T_\rho P_1 \hookrightarrow T_\rho P_2 \) and \( di_\rho^* : T_\rho^* P_2 \to T_\rho^* P_1 \). By identifying the dual \( T_\rho^* P_i \) with \( T_\rho P_i \) \( (i = 1, 2) \), we may regard \( di_\rho^* \) as \( di_\rho^* : T_\rho P_2 \to T_\rho P_1 \). Let \( g \) be a nonnegative inner product on \( T_\rho P_1 \), then \( di_\rho g di_\rho^* \) is a nonnegative inner product on \( T_\rho P_2 \).

Lemma 15: We have the inequality

\[
\max_{(a,S) \in U^*(g)} (\text{tr}_{T_\rho P_1} a + \text{tr}_H S) \leq \max_{(a',S) \in U^*(di_\rho g di_\rho^*)} (\text{tr}_{T_\rho P_2} a' + \text{tr}_H S). \tag{19}
\]

Moreover the equality holds, if and only if there exists \( (a',S) \in U^*(di_\rho g di_\rho^*) \) such that \( a'(T_\rho P_1) \subset T_\rho P_1 \), and the maximum of the right hand side is attained by \( (a',S) \).

Proof: We have \( (di_\rho a di_\rho^*,S) \in U^*(di_\rho g di_\rho^*) \) for \( (a,S) \in U^*(g) \).

\[
F : U^*(g) \to U^*(di_\rho g di_\rho^*) \quad \psi \quad \psi \quad (a,S) \mapsto (di_\rho a di_\rho^*,S). \tag{20}
\]

Then, we have \( \text{Tr}(a,S) = \text{Tr}(F(a,S)) \). In (19) the equality holds, if and only if

\[
\max_{(a',S) \in U^*(di_\rho g di_\rho^*)} \text{Tr}(a',S) = \max_{(a',S) \in \text{Im} F} \text{Tr}(a',S). \tag{21}
\]

By the definition of \( U^*(di_\rho g di_\rho^*) \), as \( di_\rho g di_\rho^*(\text{Ker } di_\rho^*) = 0 \), we have \( a'(\text{Ker } di_\rho^*) = 0 \) for \( (a',S) \in U^*(di_\rho g di_\rho^*) \). Thus, \( (a',S) \in \text{Im} F \) for \( (a',S) \in U^*(di_\rho g di_\rho^*) \), if and only if \( a'(T_\rho P_1) \subset T_\rho P_1 \). Thus, the proof is complete.

5. 3-Parameter Spin-1/2 Model

In this section, we consider a 3-parameter spin-1/2 model. Let us define the Pauli matrices \( \sigma_1, \sigma_2, \sigma_3 \) in the usual way:

\[
\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\]
Assume that \( \dim \mathcal{H} = 2 \), \( \dim T_\rho P = 3 \), and that \( \rho \in T_{sa}^{+1}(\mathcal{H}) \) is not a pure state. We may assume that \( \rho = \frac{1}{2}(\text{Id} + \alpha \sigma_3) \). By \( S(T_\rho P) \) we denote the unit sphere in \( T_\rho P \) with respect to the *SLD inner product. We assume that \( g \) is a quadratic form on \( T_\rho P \). In this section, we assume \( T_\rho P \) to be a real Hilbert space with respect to the *SLD inner product \( J_{S,T}^\rho \). \( f^3 = \sqrt{1 - \alpha^2} \sigma_3 \), \( f^i = \frac{\sigma^i}{2} \) (\( i = 1, 2 \)) is an orthonormal base on \( T_\rho P \). The dual base of \( f^i \) is \( f_3 = \frac{-\alpha}{\sqrt{1 - \alpha^2}} \text{Id} + \frac{1}{\sqrt{1 - \alpha^2}} \sigma_3 \), \( f_i = \sigma_i \) \( (i = 1, 2) \). We need some lemmas.

**Lemma 16:** For \( y \in T_\rho P \), we have

\[
\det(r \text{Id}_\mathcal{H} - J^{-1}(y)) = \left( r + \frac{\alpha y^3}{\sqrt{1 - \alpha^2}} + \sqrt{\frac{\alpha^2(y^3)^2}{(1 - \alpha^2)} + 1} \right) \left( r + \frac{\alpha y^3}{\sqrt{1 - \alpha^2}} - \sqrt{\frac{\alpha^2(y^3)^2}{(1 - \alpha^2)} + 1} \right).
\]

**Proof:** We have

\[
J^{-1}(y) = y^3 \frac{1}{1 - \alpha^2} (-\alpha \text{Id}_\mathcal{H} + \sigma_3) + \sum_{i=2}^{3} y^i \sigma_i.
\]

Since there exists \( t \in \mathbb{R} \) s.t. \( \exp(\sqrt{-1}t \sigma_3)(y^1 \sigma_1 + y^2 \sigma_2) \exp(-\sqrt{-1}t \sigma_3) = \sqrt{(y^1)^2 + (y^2)^2} \sigma_1 \), we may assume that \( y^2 = 0 \). We have

\[
J^{-1}(y) = \begin{pmatrix}
\frac{-\alpha + 1}{\sqrt{1 - \alpha^2}} y^3 & y^1 \\
y^1 & \frac{-\alpha - 1}{\sqrt{1 - \alpha^2}} y^3
\end{pmatrix}.
\]

Therefore,

\[
\det(r \text{Id}_\mathcal{H} - J^{-1}(y)) = r^2 + \frac{\alpha y^3}{\sqrt{1 - \alpha^2}} r - (y^1)^2 - (y^3)^2
\]

\[
= \left( r + \frac{\alpha y^3}{\sqrt{1 - \alpha^2}} + \sqrt{\frac{\alpha^2(y^3)^2}{(1 - \alpha^2)} + 1} \right) \left( r + \frac{\alpha y^3}{\sqrt{1 - \alpha^2}} - \sqrt{\frac{\alpha^2(y^3)^2}{(1 - \alpha^2)} + 1} \right).
\]

**Lemma 17:** For \( z \in S(T_\rho P) \), we have

\[
J^{-1}(z) \cdot \rho \cdot J^{-1}(z) = (\text{Id}_\mathcal{H} - \rho).
\]

(23)
Proof: In the same way as above, we may assume that $z^3 = 0$. Then we have
\[
J^{-1}(z) \cdot \rho \cdot J^{-1}(z) = \begin{pmatrix}
\frac{-1+\alpha}{\sqrt{1-\alpha^2}} z^3 & \frac{1+\alpha}{\sqrt{1-\alpha^2}} z^3 \\
\frac{-1-\alpha}{\sqrt{1-\alpha^2}} z^3 & \frac{1-\alpha}{\sqrt{1-\alpha^2}} z^3
\end{pmatrix} \begin{pmatrix}
\frac{1+\alpha}{2} & 0 \\
0 & \frac{1-\alpha}{2}
\end{pmatrix} 
\begin{pmatrix}
\frac{-1+\alpha}{\sqrt{1-\alpha^2}} z^3 & \frac{1+\alpha}{\sqrt{1-\alpha^2}} z^3 \\
\frac{-1-\alpha}{\sqrt{1-\alpha^2}} z^3 & \frac{1-\alpha}{\sqrt{1-\alpha^2}} z^3
\end{pmatrix} 
\]
\[
= \begin{pmatrix}
\frac{1-\alpha}{2} & 0 \\
0 & \frac{1+\alpha}{2}
\end{pmatrix} = \text{Id}_H - \rho.
\]
\[
\]
For $x \in T_\rho P$, we denote the spectral measure of $J^{-1}x$ by $M_{J^{-1}x} \in \mathcal{M}_s(\mathbb{R}, \mathcal{H})$. For $y \in T_\rho P$ we define the map $(y)$ in the following way:
\[
(y) : \mathbb{R} \to T_\rho P \\
\psi \mapsto (y) \\
c \mapsto c \cdot y.
\] (24)
Put $M^y_{J^{-1}x} := (y)^* M_{J^{-1}x} \in \mathcal{M}(T_\rho P, \mathcal{H})$. This is important to establish the optimal measurement.

**Lemma 18:** We have the following equation with respect to $E(M^y_{J^{-1}x}) \in \text{Hom}(\mathcal{L}^2_{sa}(\rho), T_\rho P) \subset \text{Hom}(\mathcal{L}^2_{sa}(\rho), \mathcal{L}^2_{sa}(\rho))$.
\[
E(M^y_{J^{-1}x}) = y \otimes J^{-1}x = |y\rangle \langle x|.
\] (25)
Proof: For $\tau \in \mathcal{L}^2_{sa}(\rho) = \mathcal{T}_{sa}(\mathcal{H})$, we have
\[
E(M^y_{J^{-1}x})(\tau) = \int_{T_\rho P} x_1 \text{tr}_\mathcal{H} M^y_{J^{-1}x} (dx_1) \tau = y \text{tr}_\mathcal{H} \tau J^{-1}x.
\] (26)

By Theorem 10 and Corollary 11 we obtain the following theorem about the attainable Cramér-Rao type bound.

**Theorem 19:** Let $g$ be a nonnegative inner product on $T_\rho P$, then
\[
\inf_{M \in \mathcal{U}(T_\rho P)} \mathcal{D}^p_g(M) = (\text{tr}_{T_\rho P} \sqrt{g})^2.
\] (27)
Moreover, the optimal measurement is given by a random measurement (i.e., a convex combination of simple measurements).
Proof: Take the Lagrange multipliers in the following way:

\[ a(x) := 2\beta \cdot \sqrt{Jg}x, \quad S := -\frac{\beta^2}{2}(\text{Id} - \alpha\sigma_3) = -\beta^2(\text{Id}\mathcal{H} - \rho), \]

where we put \( \beta := \text{tr}_{T_P}\sqrt{Jg} \). Then, we have

\[
R^\rho_g(a, S; x) = g(x, x) \cdot \rho - a(x) - S \\
= J^{-1}(\sqrt{Jg}x, \sqrt{Jg}x) \cdot \rho - 2\beta \cdot \sqrt{Jg}x + \beta^2(\text{Id}\mathcal{H} - \rho) \\
= J^{-1}(y, y) \cdot \rho - 2\beta \cdot y + \beta^2(\text{Id}\mathcal{H} - \rho) = r^2 \rho - 2\beta r + \beta^2(\text{Id}\mathcal{H} - \rho) \\
= r^2 \rho - 2\beta r \circ J^{-1}(z) + \beta^2(\text{Id}\mathcal{H} - \rho) \\
= (r - \beta J^{-1}(z))\rho(r - \beta J^{-1}(z)) + \beta^2((\text{Id}\mathcal{H} - \rho) - J^{-1}(z) \cdot \rho \cdot J^{-1}(z)) \\
= (r - \beta J^{-1}(z))\rho(r - \beta J^{-1}(z)),
\]

where \( z \in S_{J^{-1}}(T_P P), \ r > 0, \ y = r \cdot z, \ y = \sqrt{Jg}(x) \). We derive the last equation from Lemma 17. Therefore, \((a, S) \in \mathcal{U}^*(g)\). Thus, \( \text{tr}_{T_P} \rho a + \text{tr}_{\mathcal{H}} S = \left(\text{tr}_{T_P} \sqrt{Jg}\right)^2 \) is a Cramér-Rao type bound.

First let us consider the case in which \( g \) is nondegenerate. By Lemma 16, one of eigenvalues of \( \beta J^{-1}z \) is positive and another is negative. We denote the positive one by \( r(z)^+ \), another by \( r(z)^- \). We denote their eigenvectors by \( v(z)^+, \ v(z)^- \), respectively. As we have \( r(-r)^+ = -r(z)^-, \ r(-z)^- = -r(z)^+, \ v(-z)^+ = v(z)^-, \ v(-z)^- = v(z)^+ \), we have

\[
\mathcal{R}^\rho_g(a, S; M^{\sqrt{Jg}^{-1}z}) \\
= \left\langle v(z)^+ \bigg| (r(z)^+ - \beta J^{-1}(z))\rho(r(z)^+ - \beta J^{-1}(z)) \bigg| v(z)^+ \right\rangle \\
+ \left\langle v(-z)^+ \bigg| (r(-z)^+ - \beta J^{-1}(z))\rho(r(-z)^+ - \beta J^{-1}(z)) \bigg| v(-z)^+ \right\rangle \\
= 0. \tag{28}
\]

Diagonalize \( \sqrt{Jg} \) in the following way:

\[
\sqrt{Jg} = \sum_{i=1}^3 W_i e_i \otimes J^{-1}e_i = \sum_{i=1}^3 W_i |e_i\rangle\langle e_i|, \ e_i \in S(T_P P). \tag{29}
\]

Put \( M := \sum_{i=1}^3 \frac{W_i}{\beta} M^{\sqrt{Jg}^{-1}e_i} \in \mathcal{M}(T_P\mathcal{H}) \). Then, we have \( \mathcal{R}^\rho_g(a, S; M) = \)
0. By (7) and Lemma 18, we have
\[
E(M) = \sum_{i=1}^{3} \frac{W_i}{\beta} E(M^{\sqrt{Jg^{-1}}e_i}) = \sum_{i=1}^{3} \frac{W_i}{\beta} \sqrt{Jg^{-1}} e_i \langle \beta e_i \rangle
\]
\[
= \sum_{i=1}^{3} \frac{W_i}{\beta} |W_i^{-1} e_i \rangle \langle \beta e_i | = \sum_{i=1}^{3} |e_i \rangle \langle e_i |.
\]

Thus, the measurement \( M \in \mathcal{M}(T_\rho P, \mathcal{H}) \) satisfies the locally unbiasedness conditions. As it satisfies the conditions of Corollary 3.1, the random measurement \( M \in \mathcal{M}(T_\rho P, \mathcal{H}) \) attains a Cramér-Rao type bound \((\text{tr}_{T_\rho P} \sqrt{Jg})^2\).

Next let us consider the case in which \( g \) is degenerate. Let \( \{A_n\} \) be a sequence of positive valued \( 3 \times 3 \) symmetric matrices with \( \lim_{n \to \infty} A_n = Jg \), where every \( A_n \) and \( \sqrt{Jg} \) commute with each other, and \( \text{tr}_\mathcal{H} A_n = \sqrt{Jg} = \beta \). Let us take complete orthonormal system \( e_1, e_2, e_3 \) of the eigenvectors of \( A_n, Jg \). Let \( W_i^n, i = 1, 2, 3 \) be the eigenvalues of \( \sqrt{A_n} \). We define \( M_n \in \mathcal{M}(T_\rho P, \mathcal{H}) \) in the following way:
\[
M_n := \sum_{i=1}^{3} \frac{W_i^n}{\beta_n} M_{\beta J^{-1} e_i}.
\] (30)

We can show \( M_n \in \mathcal{U}(T_\rho P) \) in the same way as the case in which \( g \) is nondegenerate. Then, we have
\[
\lim_{n \to \infty} \mathcal{R}^\rho_g (a, S; M_n) = \sum_{i=1}^{3} \lim_{n \to \infty} \frac{W_i^n}{\beta} \mathcal{R}^\rho_g (M_{\beta J^{-1} e_i}^A) = 0.
\]

The proof is complete.

Finally let us consider a 2-parameter case. Put \( P_1 := T_{sa}^{+, 1}(C^2) \) and assume that \( P_2 \subset P_1, \dim P_2 = 2 \). Let \( g := \sum_{i=1}^{2} (W_i)^2 e_i \otimes e_i \) be a quadratic form on \( T_\rho P_2 \), where \( e_1, e_2 \) is complete orthonormal system of \( T_\rho^* P_2 \). The maximum of the duality problem is given the following Lagrange multiplier,
\[
a(x) := 2\beta \cdot (\sqrt{Jg})(x), \quad S := -\beta^2 (Id - \rho), \quad \beta = \text{tr} \sqrt{Jg},
\]

\((a, S)\) satisfies the conditions of the equality of Lemma 15. The optimal measurement is given by the random measurement which is the limit of the optimal measurement sequence given in the case of degenerate \( g \) in Theorem 19.
Acknowledgments

I wish to thank Professor A. Fujiwara for introducing me into this subject, and Professor K. Ueno for useful comments about this paper.

References


CHAPTER 13
Statistical Model with Measurement Degree of Freedom and Quantum Physics

Masahito Hayashi and Keiji Matsumoto

Abstract. The asymptotic efficiency of statistical estimate of unknown quantum states is discussed, both in adaptive and collective settings. Adaptive bounds are written in single-letterized form, and collective bounds are written in limiting expression. Our arguments clarify mathematical regularity conditions.

1. Introduction

In the estimation of the unknown density operator by use of the experimental data, the error can be reduced by the improvement of the design of the experiment. Therefore, it is natural to ask what is the limit of the improvement. To answer the question, Helstrom [2] founded the quantum estimation theory, in analogy with classical estimation theory (in the manuscript, we refer to statistical estimation theory of probability distribution as ‘classical estimation theory’). Often, for simplicity, it is assumed that a state belongs to a family $\mathcal{M} = \{\rho_\theta | \theta \in \Theta \subset \mathbb{R}^m\}$ of states, which is called model and that the finite dimensional parameter $\theta$ is to be estimated statistically.

He considered the quantum analogue of Cramér-Rao inequality, which gives the lower bound of mean square error of locally unbiased estimate. This bound, however, is not achievable at all, when the number of the data is finite.

Let us assume that the number $n$ of the data tends to infinite. Then, if some regularity conditions are assumed, it is concluded that if the estimate is consistent, i.e., the estimate converges to the true value of parameter, the first order asymptotic term of mean square error satisfies the Cramér-Rao inequality, and that the bound is achieved for all $\theta \in \Theta$. This kind of discussion is called first order asymptotic theory.

The quantum version of first order asymptotic theory is started by H. Nagaoka [6,7]. He defined, in our terminology, the quasi-quantum
Cramér-Rao type bound, and pointed out that the bound is achieved asymptotically and globally. The proof of achievability, however, is only roughly sketched in his paper. In this manuscript, the proof of the achievability of the bound is fully written out, and the regularity conditions for the achievability is revealed. In addition, we defined another bound, the quantum Cramér-Rao type bound, and showed that the new bound is also achievable, if the use of quantum correlation between samples are allowed.

2. Preliminaries

An estimate \( \hat{\theta} \) is obtained as a function \( \hat{\theta}(\omega) \) of data \( \omega \in \Omega \) to \( \mathbb{R}^m \). The purpose of the theory is to obtain the best estimate and its accuracy. The optimization is done by the appropriate choice of the measuring apparatus and the function \( \hat{\theta}(\omega) \) from data to the estimate.

Let \( \sigma(\mathbb{R}^m) \) be a \( \sigma \)-field in the space \( \mathbb{R}^m \). Whatever apparatus is used, the data \( \omega \in \Omega \) lie in a measurable subset \( B \in \sigma(\mathbb{R}^m) \) of \( \Omega \) writes

\[
\Pr\{\omega \in B|\theta\} = \text{tr} \rho(\theta) M(B),
\]

when the true value of the parameter is \( \theta \). Here, \( M \), which is called positive operator-valued measure (POM, in short), is a mapping from subsets \( B \subset \Omega \) to non-negative Hermitian operators in \( \mathcal{H} \), such that

\[
M(\phi) = O, \quad M(\Omega) = I, \quad M\left( \bigcup_{i=1}^{\infty} B_i \right) = \sum_{i=1}^{\infty} M(B_i) \quad (B_i \cap B_j = \phi, \ i \neq j)
\]

(see p. 53 [2] and p. 50 [3]). Conversely, some apparatus corresponds to any POM \( M \) [8, 9]. Therefore, we refer to the measurement which is controlled by the POM \( M \) as ‘measurement \( M \)’. A pair \( (\hat{\theta}, M) \) is called an estimator.

The classical Fisher information matrix \( J^M_\theta \) by the POM \( M \) is defined, as in classical estimation theory,

\[
J^M_\theta := \left[ \int_{\omega \in \Omega} \partial_i \log \frac{dP^M_\theta}{d\nu} \partial_j \log \frac{dP^M_\theta}{d\nu} dP \right],
\]

where \( \partial_i = \partial/\partial \theta^i \), \( P^M_\theta(B) := \text{tr} \rho_\theta M(B) \), and \( \nu \) is some underlying measure (in the manuscript, we assume that for any POM \( M \), there is a measure \( \nu \) in \( \Omega \) such that \( P^M_\theta \prec \nu \) for all \( \theta \in \Theta \)). Denote the mean square error matrix of \( (\hat{\theta}, M) \) by \( V_\theta[\hat{\theta}, M] \), and, as the measure of accuracy, let us take \( \text{Tr} G V_\theta[\hat{\theta}, M] \), where \( G \) is nonnegative symmetric real matrix. If \( G = \text{diag}(g_1, \cdots, g_m) \), \( \text{Tr} G V_\theta[\hat{\theta}, M] \) is the weighed sum of mean square error of the estimate \( \hat{\theta}^i \) of each component \( \theta^i \) of the parameter.
Let us define locally unbiased estimator \((\hat{\theta}, M)\) at \(\theta\) by,
\[
E_{\theta}[\hat{\theta}, M] := \int \hat{\theta}^j(\omega) \text{tr}_{\theta} M(d\omega) = \theta^j, \quad (j = 1, \cdots, m).
\]
(3)
\[
\int \hat{\theta}^j(\omega) \text{tr}_{k\theta} M(d\omega) = \delta^j_k, \quad (j, k = 1, \cdots, m).
\]
(4)
Then, \(J^M_\theta\) is characterized by,
\[
J^M_\theta = \inf\{V_{\theta}[\hat{\theta}, M] \mid \hat{\theta} : (\hat{\theta}, M) \text{ is locally unbiased}\},
\]
and the quasi-quantum Cramér-Rao type bound \(C_\theta(G)\) is defined by,
\[
C_\theta(G) := \inf\{\text{Tr} G V_{\theta}[\hat{\theta}, M] \mid \theta \text{ is locally unbiased}\}
= \inf\{\text{Tr} G J^{M-1}_\theta \mid \theta \text{ is a POM in } \mathcal{H}\}.
\]
Nagaoka pointed out that the quasi-quantum Cramér-Rao type bound is achievable asymptotically for every \(\theta \in \Theta\) [6]. \(C_\theta(G)\) is calculated explicitly for several special cases [1, 5].

Suppose \(n\)-i.i.d. state \(\rho_{\otimes n}\) of the unknown state \(\rho_{\theta}\) are given. Then the sequence \(\{(\hat{\theta}_n, M_n)\}\), where \(M_n\) is a POM in \(\mathcal{H}_{\otimes n}\), is said to be \textit{MSE consistent} if the estimate \(\hat{\theta}_n\) converges to the true parameter in the mean square error, i.e., \(\lim_{n \to \infty} V_{\theta}[\hat{\theta}_n, M_n] = 0\).

3. The Quasi-Classical Cramér-Rao Type Bound

3.1. The lower bound
Let \(M_{(1)}, \ldots, M_{(n)}\) be a sequence of the POMs in \(\mathcal{H}\), and apply the measurement \(M_{(1)}\) to the first sample, and the measurement \(M_{(2)}\) to the second sample, and so on. The choice of \(M_{(k)}\) is dependent on the outcome \(\bar{\omega}_{k-1} = (\omega_{(1)}, \cdots, \omega_{(k-1)})\) of \(M_{(1)}, \ldots, M_{(k-1)}\). To reveal the dependency of \(M_{(k)}\) on \(\bar{\omega}_{k-1}\), we write \(M_{(k)}[\bar{\omega}_{k-1}]\).

Let us define the POM \(M_n\) in \(\mathcal{H}^n\) which takes value in \(\Omega^n\) by,
\[
M_n(B) = \int_{\bar{\omega}_n \in B} \bigotimes_{k=1}^n M_{(k)}[\bar{\omega}_{k-1}](d\omega_{(k)}).
\]
Then the data \(\bar{\omega}_n\) is controlled by the probability distribution \(P_{\theta}^{M_n}(B) = \text{tr} \rho_{\otimes n}^{\otimes n} M_n(B)\).

The estimator is said to be \textit{asymptotically unbiased} if
\[
(B_{n,\theta})^i = (B_\theta(\hat{\theta}_n, M_n))^i := \int_{\Omega} (\hat{\theta}_{n}(\omega) - \theta^i) P_{\theta}^{M_n}(d\omega) \to 0 \text{ as } n \to \infty \quad (5)
\]
\[
(A_{n,\theta})_j = (A_\theta(\hat{\theta}_n, M_n))_j := \frac{\partial}{\partial \theta^j} E_{\theta}^i[\hat{\theta}_n, M_n] \to \delta_j^i \text{ as } n \to \infty. \quad (6)
\]
The MSE consistent estimator satisfies (5) always. Therefore, if appropriate regularity conditions are assumed so that the differential, the integral and the trace commute with each other, then (6) is also satisfied, and the estimator will be asymptotically unbiased.

**Theorem 1:** If \( \{(\hat{\theta}_n, M_n)\} \) is MSE consistent, and \( \lim_{n \to \infty} n V_{\theta}[(\hat{\theta}_n, M_n)] \) exists,

\[
\lim_{n \to \infty} n \text{Tr} G V_{\theta}[(\hat{\theta}_n, M_n)] \geq C_{\theta}(G). \tag{7}
\]

**Proof:** In the almost same manner as classical estimation theory, (6) leads to,

\[
n V_{\theta}[\hat{\theta}_n, M_n] \geq n A_n \left( J_{\theta}^{M_n} \right)^{-1} t A_n \tag{8}
\]

Elementary calculation leads to,

\[
\frac{1}{n} J_{\theta}^{M_n} = J_{\theta}^{M^n_{\theta}}, \tag{9}
\]

where \( M^n_{\theta} \in \mathcal{M} \) is a POM in \( \mathcal{H} \) which is defined by

\[
M^n_{\theta}(\prod_{k=1}^{n} B_k) = \int \sum_{k=1}^{n} M_{(k)}[\bar{\omega}_{k-1}](B_k)P_{\theta}^{M_n}(d\bar{\omega}_n).
\]

(8) and (9) yield

\[
\text{Tr} G n V_{\theta}[\hat{\theta}_n, M_n] \geq \text{Tr} G A_{n,\theta} \left( J_{\theta}^{M^n_{\theta}} \right)^{-1} t A_{n,\theta} \geq C_{\theta}(t A_{n,\theta} G A_{n,\theta}). \tag{10}
\]

Passing both sides of (10) to the limit \( n \to \infty \), we have the theorem. \( \blacksquare \)

3.2. **Estimator which achieves the bound**

The estimator defined in the following achieves the equality in the inequality (7) if the regularity conditions (B.1–4) are satisfied. The proof will be presented later in the subsection 3.4.

First, apply the measurement \( M_0 \) to \( \sqrt{n} \) samples of unknown state \( \rho_{\theta} \), and calculate \( \hat{\theta}_n \) which satisfies (12). Second, apply the measurement \( M_{\hat{\theta}_n} \) to the remaining \( n - \sqrt{n} \) samples, where \( M_{\theta} \) is defined by

\[
\text{Tr} G \left( J_{\theta}^{M_{\theta}} \right)^{-1} \leq C_{\theta}(G) + \varepsilon', \tag{11}
\]
(11) is satisfied. Then, $\hat{\theta}_n$ is defined to be $\overline{\theta}_n(\hat{\theta}_n)$, where $\overline{\theta}_n(\theta')$ is defined by,

$$\overline{\theta}_n(\theta') = \arg\max_{\theta \in \Theta} \sum_{k=\sqrt{n}+1}^{n} \log \frac{dP_{M_{\theta'}}}{d\nu}(\omega_k).$$

3.3. Regularity conditions

(B.1) There is a POM $M_0$ and $\hat{\theta}_n$ which satisfies

$$\lim_{n \to \infty} P_{\hat{\theta}_n}^{M_n} \{ \| \theta - \hat{\theta}_n \| > \delta \} = 0, \ \forall \delta > 0. \quad (12)$$

(B.2) $K := \sup_{\theta \in \Theta} \| \theta \|$ is finite.

(B.3) $\overline{\theta}_n(\theta')$ achieves the equality in the classical asymptotic Cramér-Rao inequality of the family $\{P_{M_{\theta'}} | \theta \in \Theta\}$ of probability distributions.

(B.4) The higher order term of mean square error of $\overline{\theta}_n(\theta')$ is uniformly bounded when $\| \theta' - \theta \| < \delta_1$ for some $\delta_1 > 0$. In other words, for any $\varepsilon > 0, \theta \in \Theta$, there exists a positive real number $\delta_1 > 0$ and a natural number $N$ such that,

$$\left| (n - \sqrt{n}) \text{Tr} G V_{\theta,n} - \text{Tr} G \left( J_{\theta'}^{-1} \right) \right| < \varepsilon, \quad (13)$$

$$\forall n \geq N, \ \forall \theta' \text{ s.t. } \| \theta - \theta' \| \leq \delta_1,$$

where $V_{\theta,n}$ is the conditional mean square error matrix of $\overline{\theta}_n(\theta')$ when $\theta'$ is given.

(B.5) For any $\varepsilon > 0, \theta \in \Theta$, there exists $\delta_2 > 0$, such that,

$$\left| \text{Tr} G \left( J_{\theta'}^{-1} \right) - C_{\theta}(G) \right| < \varepsilon, \ \forall \theta, \forall \tilde{\theta} \text{ s.t. } \| \theta - \tilde{\theta} \| < \delta_2. \quad (14)$$

(B.1) is satisfied almost always, and (B.2) is not restrictive. For $\overline{\theta}_n(\theta')$ is the maximum likelihood estimator of the family $\{P_{M_{\theta'}} | \theta \in \Theta\}$ of probability distributions, (B.3) is satisfied in usual cases. The validity of (B.4), however, is hard to verify. Therefore, in the future, this condition needs to be replaced by other conditions. Obviously, (B.5) reduces to the following (B.5.1–2), both of which are natural.

(B.5.1) The map $\theta \mapsto C_{\theta}(G)$ is continuous.

(B.5.2) For any $\theta'$, the map $\theta \mapsto [J_{\theta'}^{-1}]$ is continuous.
3.4. Proof of achievability

**Theorem 2:** If the model $\mathcal{M}$ satisfy conditions (B.1–5) in the following, then we have,
\[
\lim_{n \to \infty} n \text{Tr} G\theta[\hat{\theta}_n, M_n] = C_\theta(G), \forall \theta \in \Theta,
\]

**Proof:** Let us choose $\delta_1, \delta_2$ and $N$ so that (13 – 14) are satisfied, and define $\delta' := \min(\delta_1, \delta_2)$. Then, if $n \geq N$, we have,
\[
\begin{align*}
&n \text{Tr} G\theta[\hat{\theta}_n, M_n] \\
= &n \int \text{Tr} G\theta[\bar{\theta}_n(\bar{\theta}_n), M_n] P^M_{\theta_n}(d\omega) \\
\leq &n \int \text{Tr} G\theta[\bar{\theta}_n(\bar{\theta}_n), M_n] P^M_{\theta_n}(d\omega) \\
&+ K^2 \text{Tr} G \int_{||\theta - \bar{\theta}_n|| > \delta'} P^M_{\theta_n}(d\omega) \\
\leq &\frac{n}{n - \sqrt{n}} \int_{||\theta - \bar{\theta}_n|| \leq \delta'} \left( \text{Tr} G \left( J_{\theta_n}^{M_{\theta_n}} \right)^{-1} + \varepsilon + \varepsilon' \right) P^M_{\theta_n}(d\omega) \\
&+ nK^2 \text{Tr} G P^M_{\theta_n} \{ ||\theta - \bar{\theta}_n|| > \delta \} \\
\leq &\frac{n}{n - \sqrt{n}} \int_{||\theta - \bar{\theta}_n|| \leq \delta} (C_\theta(G) + 2\varepsilon + \varepsilon') P^M_{\theta_n}(d\omega) \\
&+ nK^2 \text{Tr} G P^M_{\theta_n} \{ ||\theta - \bar{\theta}_n|| > \delta \} \\
\leq &\frac{n}{n - \sqrt{n}} (C_\theta(G) + 2\varepsilon + \varepsilon') + nK^2 \text{Tr} G P^M_{\theta_n} \{ ||\theta - \bar{\theta}_n|| > \delta \}.
\end{align*}
\]

(B.1) implies that the third term of the last end of the equation tends to $n$ as $n \to \infty$. Therefore, we have, for every $\varepsilon' > 0$ and for every $\varepsilon > 0$,
\[
\lim_{n \to \infty} n \text{Tr} G\theta[\hat{\theta}_n, M_n] \leq C_\theta(G) + 2\varepsilon + \varepsilon'.
\]
which leads to the theorem.  

4. Use of Quantum Correlation

In this section, we consider the minimization of asymptotic mean square error where $M_n$ runs every POM which satisfies MSE consistency. Physically, this means we allow the use of interactions between samples.

So far, we considered POM which takes value in $\Omega$, or the totality of all the possible data. Instead, in this section, we consider POM with the values in $\mathbb{R}^d$, for if $M$ is a POM with values in $\Omega$, $M \circ \hat{\theta}^{-1}$ is POM with the
values in \( \mathbb{R}^d \). MSE consistency is defined in the same way as the precedent sections.

Let \( C_n^\theta(G) \) denote the quasi-quantum Cramér-Rao type bound of the family \( \{ \rho_\theta \otimes^n | \theta \in \Theta \} \) of density operators in \( \mathcal{H}^\otimes n \). Then, the quantum Cramér-Rao type bound \( C_Q^\theta(G) \) is defined by,

\[
C_Q^\theta(G) := \lim_{n \to \infty} nC_n^\theta(G).
\]

For \( C_\theta(G) \geq nC_n^\theta(G) \) holds true, we have,

\[
C_\theta(G) \geq C_Q^\theta(G).
\]

**Theorem 3:** If the sequence \( \{ M^n \}_{n=1}^{\infty} \) is MSE consistent, we have,

\[
\lim n_1 \, \text{tr} \, G V_\theta[M_n] \geq C_Q^\theta(G).
\]  

(16)

**Proof:** In the almost same manner as the proof of theorem 1, we have,

\[
V_\theta[M_n] \geq A_{n,\theta} \left( J_{\theta}^M \right)^{-1} t A_{n,\theta},
\]

\[
n \, \text{tr} \, G V_\theta[M_n] \geq n \, \text{tr} \, G A_{n,\theta} \left( J_{\theta}^M \right)^{-1} t A_{n,\theta} \geq nC_n^\theta(t A_{n,\theta} G A_{n,\theta}),
\]

which approaches (16) as \( n \to \infty \).

If the family \( \{ \rho_\theta \otimes^n | \theta \in \Theta \} \) of density operators satisfies (B.1–5), we have the following theorem.

**Theorem 4:** There is an MSE consistent sequence \( \{ M_n \} \) of POM such that \( \lim_{n \to \infty} n \, \text{tr} \, G V_\theta[M_n] \leq C_Q^\theta(G) + \epsilon \) is satisfied for every \( \epsilon > 0 \) and for every \( \theta \in \Theta \).

**Proof:** Let us divide \( n \) samples into \( n_2 \) groups each of which is consist of \( n_1 \) samples, and let \( M_{(1)}^{n_1}, \ldots, M_{(n_2)}^{n_1} \) be a sequence of POMs in \( \mathcal{H}^\otimes n_1 \). Apply the measurement \( M_{(1)}^{n_1} \) to the first group \( \rho_\theta \otimes^{n_1} \) of samples, and apply \( M_{(2)}^{n_1} \) to the second samples, and so on. The choice of \( M_{(k)}^{n_1} \) is dependent on the outcome of the measurements \( M_{(1)}^{n_1}, \ldots, M_{(k-1)}^{n_1} \). With \( n_1 \) fixed, let us approach \( n_2 \) to \( \infty \). Then, theorem 2 implies the existence of a MSE consistent sequence \( \{ M_n \} \) of POM which satisfies

\[
\lim n_1 \, \text{tr} \, G V_\theta[M_n] = \lim n_1 n_2 \, \text{tr} \, G V_\theta[M_n] = n_1 C_{n_1}^\theta(G).
\]  

(17)

For any epsilon, if \( n_1 \) is sufficiently large, \( \lim_{n \to \infty} n \, \text{tr} \, G V_\theta(M^n) \leq C_Q^\theta(G) + \epsilon \) is satisfied, and we have the theorem.
References


CHAPTER 14

Asymptotic Quantum Theory for the Thermal States Family

Masahito Hayashi

Abstract. Concerning state estimation, we will compare two cases. In one case we cannot use the quantum correlations between samples. In the other case, we can use them. In addition, under the later case, we will propose a method which simultaneously measures the complex amplitude and the expected photon number for the displaced thermal states.

1. Introduction

Quantum estimation is essentially different from classical estimation regarding the following two points. The first point is that we cannot simultaneously construct the optimal estimators corresponding to respective parameters because of non-commutativity between them. It has been a serious problem since the beginning of the quantum estimation [4, 5, 8]. The second point is that we can reduce the estimation error under the assumption that we can prepare independent and identical samples of the unknown quantum state. It was pointed by Nagaoka [6, 7] concerning the large deviation theory in one-parameter estimation. The purpose of this paper is to clear the second point concerning the mean square error (MSE).

Our situation is divided into the following two cases. In the first case, we estimate the unknown state by independently measuring every sample. In this case, we may decide the \( n \)-th POVM from \( n - 1 \) data which have been already given. In the second case, we estimate the unknown state by regarding \( n \)-sample system as a single composite system. In this case, we may use POVMs which are indivisible into every sample system. In order to construct these POVMs, we need to use quantum correlations between every sample. The former is called the non-quantum correlation case and
the later the quantum correlation case. When the unknown state is a pure
state, the errors of both are asymptotically equivalent in the first order [2].
Concerning the spin 1/2 system, see Hayashi [1].

In this paper, we formulate a general theory for the asymptotic quantum
estimation. It is applied to the simultaneous estimation of the expected pho-
ton number and the complex amplitude for the quantum displaced thermal
state.

2. Asymptotic Estimation Theory

In this paper, we use a quantum state family $S$ parameterized by finite
parameters $\theta^1, \ldots, \theta^d$:

$$S := \{ \rho_\theta \in S(\mathcal{H}) | \theta = (\theta^1, \ldots, \theta^d) \in \Theta \subset \mathbb{R}^d \},$$

where the set $S(\mathcal{H})$ denotes the set of densities on $\mathcal{H}$. For simplicity, we
assume that $\rho_\theta$ is nondegenerate.

2.1. Non-quantum correlation case

The non-quantum correlation case is formulated as follows. A pair $E_n = (\{M_k\}_{k=1}^n, \hat{\theta}_n)$ is called a recursive estimator where $\hat{\theta}_n$ is a function esti-
mating the unknown parameter from $n$ data, and $\{M_k\}_{k=1}^n$ is a sequence
of POVMs $M_1, M_2(\omega_1), \ldots, M_n(\omega_1, \ldots, \omega_{n-1})$ as follows: the $n$-th POVM
$M_k(\omega_1, \ldots, \omega_{k-1})$ is determined by $k-1$ data which have been already
given. A sequence $\{E_n\}_{n=1}^\infty$ of recursive estimators is called a recursive MSE
consistent estimator if

$$\lim_{n \to \infty} \frac{1}{n} \sum_{n} \| \hat{\theta}_n(\omega_1, \omega_2, \ldots, \omega_n) - \theta \|^2 = 0,$$

for all $\theta \in \Theta$.

where

$$P_{\rho}^{E_n}(d\omega_1, d\omega_2, \ldots, d\omega_n) = \text{tr} \rho M_1(d\omega_1) \text{tr} \rho M_2(\omega_1)(d\omega_2) \ldots$$

$$\ldots \text{tr} \rho M_n(\omega_1, \ldots, \omega_{n-1})(d\omega_n).$$

We define the non-quantum-correlational Cramér-Rao type bound $C_\theta^{NQC}(G)$
for a weighted matrix $G$ ($G$ is a $d \times d$ real positive symmetric matrix.) as:

$$C_\theta^{NQC}(G) := \inf \left\{ \lim_{n \to \infty} \text{Tr} nGV_\theta(E_n) \right\} \{E_n\}_{n=1}^\infty \text{ is a recursive MSE consistent estimator} \right\},$$
where the MSE matrix $V_\theta (\mathcal{E}_n)$ is given by:

$$V^{i,j}_\theta (\mathcal{E}_n) = \int \cdots \int \left( \hat{\theta}^i_n (\omega_1, \ldots, \omega_n) - \theta^i \right) \left( \hat{\theta}^j_n (\omega_1, \ldots, \omega_n) - \theta^j \right) \times P^\mathcal{E}_n (d\omega_1, \ldots, d\omega_n).$$

We have the following equation:

$$C_{NQC}^\theta (G) = \inf \left\{ \text{Tr} \, G \left( J_M^\theta \right)^{-1} \middle| \text{M is a POVM on } \mathcal{H} \right\},$$

where $J_M^\theta$ denotes the Fisher information matrix at $\theta$ of $\{ \text{tr} \, \rho_M (d\omega) | \theta \in \Theta \}$. It is derived by Jensen’s inequality [3]. Under some regularity conditions, we show that there exists a recursive MSE consistent estimator $\{ \mathcal{E}_n \}_{n=1}^\infty$ such that [3]:

$$n \text{Tr} \, GV_\theta (\mathcal{E}_n) \to C_{NQC}^\theta (G) \text{ as } n \to \infty, \forall \theta \in \Theta.$$

According Holevo [5], we have $J_M^\theta \leq \tilde{J}_\theta$, where $\tilde{J}_\theta$ is the RLD Fisher information matrix defined as: $\tilde{J}_{\theta; i,j} := \text{tr} (\tilde{L}_{\theta; i})^* \rho_\theta \tilde{L}_{\theta; j}$, $\tilde{L}_{\theta; i} := (\rho_\theta)^{-1} \frac{\partial \rho_\theta}{\partial \theta^i}$.

Therefore we have the following inequality

$$C_{NQC}^\theta (G) \geq C_R^\theta (G),$$

where

$$C_R^\theta (G) := \inf \left\{ \text{tr} \, GV | V \text{ is a } d \times d \text{ real symmetric matrix } V \geq \tilde{J}_\theta^{-1} \right\}$$

$$= \text{Tr} \, G \text{Re} \tilde{J}_\theta^{-1} + \text{Tr} \left| \sqrt{G} \text{Im} \tilde{J}_\theta^{-1} \sqrt{G} \right|. $$

### 2.2. Quantum correlation case

Next, we formulate the quantum correlation case. For this purpose, we consider a quantum counterpart of independent and identically distributed condition. If $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are $n$ Hilbert spaces which correspond to the physical systems, then their composite system is represented by the tensor Hilbert space.

$$\mathcal{H}^{(n)} := \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n.$$ 

Thus, a state on the composite system is denoted by a density operator $\rho^{(n)}$ on $\mathcal{H}^{(n)}$. In particular if $n$ element systems ($\mathcal{H}_1, \ldots, \mathcal{H}_n$) of the composite system $\mathcal{H}^{(n)}$ are independent of each other, there exists a density $\rho_k$ on $\mathcal{H}_k$ such that

$$\rho^{(n)} = \rho_1 \otimes \cdots \otimes \rho_n, \text{ on } \mathcal{H}^{(n)}.$$
The condition:

\[ \mathcal{H}_1 = \cdots = \mathcal{H}_n = \mathcal{H}, \quad \rho_1 = \cdots = \rho_n = \rho \]

corresponds to the independent and identically distributed condition in the classical case. Therefore, we consider the parameter estimation problem for the family \( \{\rho_\theta^{(n)}\} \) on \( \mathcal{H}[\theta \in \Theta] \) which is called the \emph{n-i.i.d. extended family}.

In this case, we use a sequence \( \{M^n\}_{n=1}^{\infty} \) of POVMs where \( M^n \) is a POVM on \( \mathcal{H} \) whose measurable set is \( \mathbb{R}^d \) as an estimator. A sequence \( \{M^n\}_{n=1}^{\infty} \) is called an MSE consistent estimator if

\[
\lim_{n \to \infty} \int_{\mathbb{R}} \|\hat{\theta} - \theta\|^2 \text{tr} \rho_\theta^{(n)} M^n (d\hat{\theta}) = 0, \quad \forall \theta \in \Theta.
\]

A recursive MSE consistent estimator can be regarded as an MSE consistent estimator because a recursive estimator \( E_n = (\{M_k\}_{k=1}^{n}, \hat{\theta}_n) \) is regarded as a POVM \( M(E_n) \) as follows:

\[
M(E_n)(B) := \int_{\theta_n^{-1}(B)} \bigotimes_{k=1}^{n} M_k(\omega_1, \ldots, \omega_{k-1}) (d\omega_k), \quad \forall B \subset \mathbb{R}^d \text{ on } \mathcal{H}^{(n)}.
\]

We define the \emph{quantum-correlational Cramér-Rao type bound} \( C_{\theta}^{QC}(G) \) for a weighted matrix \( G \) as:

\[
C_{\theta}^{QC}(G) := \inf \left\{ \liminf_{n \to \infty} n \text{Tr} G \rho_\theta^{(n)} \left| \begin{array}{c} \{M^n\}_{n=1}^{\infty} \text{ is an MSE consistent estimator} \end{array} \right. \right\}
\]

where the MSE matrix \( V_\theta(M^n) \) is given by:

\[
V_\theta^{i,j}(M^n) = \int_{\mathbb{R}^d} (\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j) \text{tr} \rho_\theta^{(n)} M^n (d\hat{\theta}).
\]

We have the following equation

\[
C_{\theta}^{QC}(G) = \liminf_{n \to \infty} n C_{\theta}^{n}(G),
\]

where \( C_{\theta}^{n}(G) \) denotes the non-quantum-correlational Cramér-Rao type bound for the \( n \)-i.i.d. extended family [3]. From the definition of the \( n \)-i.i.d. extended family, we have \( C_{\theta}^{NQC}(G) \geq nC_{\theta}^{n}(G) \). Therefore, we have the inequality \( C_{\theta}^{NQC}(G) \geq C_{\theta}^{QC}(G) \) Moreover, since \( J_{\theta}^{M^n} \leq n \tilde{J}_{\theta} \) for any POVM \( M^n \) on \( \mathcal{H}^{(n)} \), the inequality \( C_{\theta}^{QC}(G) \geq C_{\theta}^{R}(G) \) holds. Thus,

\[
C_{\theta}^{NQC}(G) \geq C_{\theta}^{QC}(G) \geq C_{\theta}^{R}(G).
\] (1)

Therefore, the difference between \( C_{\theta}^{NQC}(G) \) and \( C_{\theta}^{QC}(G) \) means the difference of the quantum correlation case from the non-quantum correlation.
case. Under some regularity conditions, we can show that there exists an
MSE consistent estimator \( \{ M^n \}_{n=1}^{\infty} \) such that \[ 3 \]:
\[
\text{Tr} \ G V_\theta (M^n) \to C_\theta^{QC} (G) \quad \text{as} \quad n \to \infty, \quad \forall \theta \in \Theta.
\]

3. Quantum Displaced Thermal States Family

Now we consider the estimation for the the complex amplitude \( \zeta \) and ex-
pected photon number \( N \) for the quantum displaced thermal states family
defined as:
\[
\mathcal{S} := \left\{ \rho_{\zeta,N} := \frac{1}{\pi N} \int_{\mathbb{C}} \exp \left( \frac{-|\zeta - \alpha|^2}{N} \right) |\alpha\rangle \langle \alpha| d^2 \alpha \ \bigg\vert \ \zeta \in \mathbb{C}, N > 0 \right\}.
\]

3.1. Estimation of complex amplitude \( \zeta \)

In the case of that photon number \( N \) is known, we estimate the tow un-
known parameters \( \zeta = (\theta^1 + i \theta^2)/\sqrt{2} \). This estimation problem is investi-
gated by Yuen & Lax [8] and Holevo [5]. In this case they calculated the
inverse \( \tilde{J}^{-1}_\theta \) of the RLD Fisher information matrix as:
\[
\tilde{J}^{-1}_\theta = \begin{pmatrix}
N + \frac{i}{2} & \frac{i}{2} \\
-\frac{i}{2} & \frac{N}{2} + \frac{1}{2}
\end{pmatrix}.
\]

They calculated the non-quantum-correlational Cramér-Rao type bound
\( C_\theta^{NQC} (G) \) as follows:
\[
C_\theta^{NQC} (G) = C_\theta^{R} (G) = 2 \left( N + \frac{1}{2} \right) g_1 + \sqrt{g_1^2 - g_2^2 - g_3^2}, \quad (2)
\]
where the weighted matrix \( G \) is parameterized as:
\[
G = \begin{pmatrix}
g_1 + g_2 & g_3 \\
g_3 & g_1 - g_2
\end{pmatrix}.
\]

From (1) and (2), we have the following equations.
\[
C_\theta^{NQC} (G) = C_\theta^{QC} (G) = C_\theta^{R} (G). \quad (3)
\]

In this case, the optimal estimator is the squeezed heterodyne.

3.2. Simultaneous estimation of complex amplitude \( \zeta \) and
expected photon number \( N \)

Next we consider the case of that both of the expected photon number \( N \)
and the complex amplitude \( \zeta \) are unknown. In this case, we estimate three
unknown parameters $\theta^1, \theta^2$ and $\theta^3 = N$. The first equation of (3) does not hold. Therefore, the squeezed heterodyne is not optimal. The inverse $\tilde{J}_{\theta}^{-1}$ of the RLD Fisher information matrix is calculated as:

$$\tilde{J}_{\theta}^{-1} = \begin{pmatrix} N + \frac{1}{2} & \frac{i}{2} & 0 \\ -\frac{i}{2} & N + \frac{1}{2} & 0 \\ 0 & 0 & N(N + 1) \end{pmatrix}.$$

Therefore we can calculate $C_{\theta}^R(G)$ as:

$$C_{\theta}^R(G) = g_0 N(N + 1) + 2 \left( N + \frac{1}{2} \right) g_1 + \sqrt{g_1^2 - g_2^2 - g_3^2},$$

if the weighted matrix $G$ can be parameterized as:

$$G = \begin{pmatrix} g_1 + g_2 & g_3 & 0 \\ g_3 & g_1 - g_2 & 0 \\ 0 & 0 & g_0 \end{pmatrix}. \quad (4)$$

If the weighted matrix $G$ can be parameterized as (4), we obtain the following equations:

$$C_{\theta}^NQC(G) > C_{\theta}^{QC}(G) = C_{\theta}^R(G). \quad (5)$$

A proof for $C_{\theta}^NQC(G) > C_{\theta}^{QC}(G)$ is omitted. The inequality $C_{\theta}^NQC(G) > C_{\theta}^{QC}(G)$ means that we cannot the simultaneous measurement of the photon number counting and heterodyne for a single sample.

3.3. Construction of an MSE consistent estimator

$\{M^n\}_{n=1}^\infty$ attaining $C_{\theta}^R(I)$

Now, for a weighted matrix $I$, we construct an MSE consistent estimator $\{M^n\}_{n=1}^\infty$ such that

$$\lim_{n \to \infty} n \operatorname{tr} V_{\theta}(M^n) = C_{\theta}^R(I), \quad \forall \theta \in \Theta.$$

It is sufficient for $C_{\theta}^{QC}(I) = C_{\theta}^R(I)$ to construct such an MSE consistent estimator.

Every POVM $M^n$ is constructed in the following step:

1. Evolve the unknown state $\rho_{\zeta,N} \otimes \cdots \otimes \rho_{\zeta,N}$ as:

$$\rho_{\zeta,N} \otimes \cdots \otimes \rho_{\zeta,N} \rightarrow U_n \rho_{\zeta,N} \otimes \cdots \otimes \rho_{\zeta,N} U_n^*$$

$$= \rho \sqrt{n_{\zeta,N}} \otimes \rho_{0,N} \otimes \cdots \otimes \rho_{0,N} \otimes \cdots \text{on } \mathcal{H}^{(n)},$$
where
\[
U_n = \exp \phi_{n-1}(a_n^* a_1 - a_1^* a_n) \cdots \exp \phi_2(a_3^* a_1 - a_1^* a_3) \exp \phi_1(a_2^* a_1 - a_1^* a_2) \text{ on } \mathcal{H}^{(n)}
\]
\[
\phi_i = \arctan \frac{1}{\sqrt{i}}, \quad i = 1, 2, \ldots, n - 1.
\]

\(a_i\) denotes the annihilation operator on \(\mathcal{H}_i\).

(2) Measure the first sample \(\rho_{\sqrt{n}, N}\) by the heterodyne, then we get the estimate \(\hat{\zeta}\) of the complex amplitude.

(3) Measure the others by the photon counting, then we obtain \(n - 1\) data which obey the probability distribution \(P^N(k)\):
\[
P^N(k) = \frac{1}{N+1} \left( \frac{N}{N+1} \right)^k, \quad k = 0, 1, \ldots
\]

(4) We obtain the estimate \(\hat{N}\) of the expected photon number \(N\) by the maximum likelihood estimator of the probability distribution \(P^N(k_1), \ldots, P^N(k_{n-1})\).

4. Conclusion

We formulate an asymptotic quantum estimation theory. This theory is applied to the simultaneous measurement of the photon number counting and the heterodyne for displaced thermal states. It is a future study to realize the MSE consistent estimator proposed in this paper in an actual physical system.

Acknowledgments

This work was supported by the Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists No. 9404. The author would like to thank Dr. K. Matsumoto for useful discussions of these topics.

References


CHAPTER 15

State Estimation for Large Ensembles

Richard D. Gill and Serge Massar

Abstract. We consider the problem of estimating the state of a large but finite number \( N \) of identical quantum systems. As \( N \) becomes large the problem simplifies dramatically. The only relevant measure of the quality of estimation becomes the mean quadratic error matrix. Here we present a bound on this quantity: a quantum Cramér-Rao inequality. This bound expresses succinctly how in the quantum case one can trade information about one parameter for information about another. The bound holds for arbitrary measurements on pure states, but only for separable measurements on mixed states—a striking example of nonlocality without entanglement for mixed but not for pure states. Cramér-Rao bounds are generally only derived for unbiased estimators. Here we give a version of our bound for biased estimators, and a simple asymptotic version for large \( N \). Finally we prove that when the unknown state belongs to a two-dimensional Hilbert space our quantum Cramér-Rao bound can always be attained and we provide an explicit measurement strategy that attains it. Thus we have a complete solution to the problem of estimating as efficiently as possible the unknown state of a large ensemble of qubits in the same pure state. The same is true for qubits in the same mixed state if one restricts oneself to separable measurements, but non-separable measurements allow dramatic increase of efficiency. Exactly how much increase is possible is a major open problem.

1. Introduction

One of the central problems of quantum measurement theory is the estimation of an unknown quantum state. Originally only of theoretical interest, this problem is becoming of increasing practical importance. Indeed there are now several beautiful experimental realizations of quantum state reconstruction in such diverse systems as quantum optics [17], molecular states [6], trapped ions [13] and atoms in motion [12].

The theoretical work which is the basis for these experiments is concerned with devising measurement strategies that are simple to realize ex-
experimentally and which allow an unambiguous reconstruction of the quantum state. The best known such technique is quantum state tomography [21], adapted in [14] for the case of finite-dimensional Hilbert spaces. However, other techniques are also available; see Ref. [22] for a recent discussion in the case of finite-dimensional Hilbert spaces. However, all these works suppose that the measurements are perfect and that any operator can be measured with infinite precision. However, in general the quality of the reconstruction will be limited by experimental error [10] or by finite statistics. The present work is devoted to studying this latter aspect, when the unknown state belongs to a finite-dimensional Hilbert space.

Thus the setting of the problem is that we may dispose of a finite number $N$ of copies of an unknown quantum state $\rho$ (pure or mixed). Our task is to determine $\rho$ as well as possible. This is by now a classical problem [8, 9].

A common approach is first to specify a cost function which numerically quantifies the deviation of the estimate from the true state. One then tries to devise a measurement and estimation strategy which minimizes the mean cost. Since the mean cost typically depends on the unknown state itself, one typically averages over all possible states to arrive at a single number expressing the quality of the estimation. However optimal strategies have only been found in some simple highly symmetric cases (the covariant measurements of Ref. [9], see also Ref. [15, 20]).

However, when the number of copies $N$ becomes large, one can hope that the problem becomes simpler so that one might be able to find the optimal strategies in this limit. The reason for this is that in the large-$N$ limit the estimation problem ceases to be a ‘global’ problem and becomes ‘local’. Indeed for small $N$ the estimated state will often be very different from the true state. Hence the optimal measurement strategy must take into account the behaviour of the cost function for large estimation errors. On the other hand, in the limit of an infinite number of copies any two states can be distinguished with certainty. So the relevant question to ask about the estimation strategy is at what rate it distinguishes neighboring states. In that case we are only concerned with the behaviour of the estimator and of the cost function very close to the true value.

To formulate the problem with precision, let us suppose that the unknown state $\rho(\theta)$ depends on a vector of $p$ unknown real parameters $\theta = (\theta_1, \ldots, \theta_p)$. For instance the $\theta_i$ could correspond to various settings or physical properties of the apparatus that produces the state $\rho$. After carrying out a measurement on the $N$ copies of $\rho$, one will guess what is $\theta$. Call $\hat{\theta}^N = (\hat{\theta}_1^N, \ldots, \hat{\theta}_p^N)$ the guessed value. For a good estimation strategy
we expect the mean quadratic error (MQE) to decrease as $1/N$:

$$E_\theta \left( (\hat{\theta}_i^N - \theta_i)(\hat{\theta}_j^N - \theta_j) \right) \equiv V_{ij}^N(\theta) \simeq \frac{W_{ij}(\theta)}{N}$$  \hspace{1cm} (1)

where the scaled MQE matrix $W(\theta) = (W_{ij}(\theta)) \simeq N V^N(\theta)$ does not depend on $N$. $E_\theta$ denotes the mean taken over repetitions of the measurement with the value of $\theta$ fixed.

Consider now a smooth cost function $f(\hat{\theta}, \theta)$, which measures how much the estimated value $\hat{\theta}$ differs from the true value $\theta$ of the parameter. $f$ will have a minimum at $\hat{\theta} = \theta$, hence can be expanded as

$$f(\hat{\theta}, \theta) = f_0(\theta) + \sum_{ij} C_{ij}(\theta)(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) + O(||\hat{\theta} - \theta||^3)$$  \hspace{1cm} (2)

where $C(\theta) = (C_{ij}(\theta))$ is a non-negative matrix. Thus for a reasonable estimation strategy the mean value of the cost will decrease as

$$E_\theta(f(\hat{\theta}^N, \theta)) = f_0(\theta) + N^{-1} \sum_{ij} C_{ij}(\theta) W_{ij}(\theta) + o(N^{-1})$$  \hspace{1cm} (3)

since we expect the expectation value of higher order terms in $\hat{\theta} - \theta$ to decrease faster than $1/N$. The problem has become local: only the quadratic cost matrix $C(\theta)$ and the scaled mean quadratic error matrix $W(\theta)$ at $\theta$ intervene. The essential question about state estimation for large ensembles is therefore what scaled MQE matrices $W(\theta)$ are attainable through arbitrary measurement and estimation procedures? In particular, what does the boundary of this set of attainable MQE matrices look like?

In the case when the parameter $\theta$ is one-dimensional ($p = 1$), the problem has been solved: a bound on the variance of unbiased estimators—the quantum Cramér-Rao bound—was given in Ref. [8], and a strategy for attaining the bound in the large-$N$ limit was proposed in Ref. [1]. This justifies taking the bound to induce a ‘distinguishability metric’ on the space of states [23, 3]. In the case of a multidimensional parameter however, though different bounds for the matrix $W$ have been established, in general they are not tight [8, 24, 9].

In this paper we present a bound for $W$ in the multiparameter case which is inspired by the discussion in [1]. This bound expresses in a natural way how one can trade information about one parameter for information about another. The interest of this new bound depends on the precise problem one is considering:

(i) When $\rho(\theta) = |\psi(\theta)\rangle\langle\psi(\theta)|$ is a pure state belonging to a 2 dimensional Hilbert space, the bound is sharp: it provides a necessary and sufficient
condition that $W$ must satisfy in order to be attainable. Furthermore, the bound can be attained by carrying out separate measurements on each particle. This completely solves the problem of estimating the state of a large ensemble of spin-1/2 particles (qubits) in the same pure state.

- **(ii)** When $\rho(\theta)$ is a pure state belonging to a Hilbert space of dimension $d$ larger than 2, then our bound on $W$ applies, but it is not sharp.

- **(iii)** When the unknown state is mixed and belongs to a 2 dimensional Hilbert space, and if one restricts oneself to measurements that act separately on each particle, then our bound applies and is sharp.

- **(iv)** When the unknown state is mixed and belongs to a Hilbert space of dimension $d > 2$, and if one restricts oneself to measurements that act separately on each particle, then our bound applies but is not sharp.

- **(v)** If the unknown state is mixed and one allows collective measurements, then our bound is not necessarily satisfied.

This last point is surprising and points to a fundamental difference between measuring pure states and mixed states. Indeed it is known that carrying out measurements on several identical copies of the same pure state can generally be done better with collective measurements on the different copies [16, 15]. This is known as “nonlocality without entanglement” [2]. The first point shows that in the limit of a large number of copies, pure states of spin-1/2 do not exhibit nonlocality without entanglement. On the other hand, the last point shows that in the limit of a large number of copies mixed states of spin-1/2 continue to exhibit nonlocality without entanglement.

To describe our bound on $W$, we first consider for simplicity the case of a pure state of spin-1/2 particles. Suppose the unknown state is a spin 1/2 known to be in a pure state, and the state is known to be almost pointing in the $+z$ direction:

$$|\psi(\theta_1, \theta_2)\rangle \simeq |\uparrow_z\rangle + \frac{1}{2}(\theta_1 + i\theta_2) |\downarrow_z\rangle$$  \hspace{1cm} (4)$$

where we have written an expression valid to first order in $\theta_1, \theta_2$. Suppose we carry out a measurement of the operator $\sigma_x$. We obtain the outcome $\pm x$ with probability $p(\pm x) = (1 \pm \theta_1)/2$. Thus the outcome of this measurement tells us about the value of $\theta_1$. Similarly we can carry out a measurement of $\sigma_y$. We obtain the outcome $\pm y$ with probability $p(\pm y) = (1 \pm \theta_2)/2$. The outcome of this measurement tells us about $\theta_2$. But the measurements $\sigma_x$ and $\sigma_y$ are incompatible, i.e., the operators do not commute and cannot be
measured simultaneously. Thus if one obtains knowledge about $\theta_1$, it is at the expense of $\theta_2$. Indeed suppose one has $N$ copies of the state $\psi$ and one measures $\sigma_x$ on $N_1$ copies and $\sigma_y$ on $N_2 = N - N_1$ copies. Our estimator for $\theta_1$ is the fraction of $+x$ outcomes minus the fraction of $-x$ outcomes. This estimator is unbiased. The resulting uncertainty (at the point $\theta_1 = \theta_2 = 0$) about $\theta_1$ is then $E_{\theta}((\hat{\theta}_1 - \theta_1)^2) = \frac{1}{N_1}$. Similarly we can estimate $\theta_2$ and the corresponding uncertainty is $E_{\theta}((\hat{\theta}_2 - \theta_2)^2) = \frac{1}{N_2}$. We can combine these two expressions in the following relation:

$$\frac{1}{E_{\theta}((\hat{\theta}_1 - \theta_1)^2)} + \frac{1}{E_{\theta}((\hat{\theta}_2 - \theta_2)^2)} = \frac{1}{V_{11}} + \frac{1}{V_{22}} = N \quad (5)$$

which expresses in a compact form how we can trade knowledge about $\theta_1$ for knowledge about $\theta_2$. We shall show that it is impossible to do better than precisely Eq. (5) when one restricts attention to unbiased estimators based on arbitrary measurements, and asymptotically not possible to do better with any estimator whatsoever.

To generalize Eq. (5), we rewrite it in a more abstract form, and state it as an inequality. We use polar coordinates to parameterize the unknown state of the spin-1/2 particle:

$$|\psi\rangle = \cos \frac{\eta}{2} |\uparrow\rangle + \sin \frac{\eta}{2} e^{i\phi} |\downarrow\rangle.$$ 

We introduce the tensor

$$H_{\eta\eta} = 1 \quad , \quad H_{\phi\phi} = \sin^2 \eta \quad , \quad H_{\eta\phi} = 0 \quad (6)$$

which is simply the Euclidean metric on the sphere. Then bound (5) can be reexpressed as

$$\text{tr} H^{-1}(V^N)^{-1} \leq N \quad (7)$$

where $V^N$ is the MQE matrix defined in Eq. (1).

For mixed states belonging to a 2 dimensional Hilbert space, (7) can be generalized as follows. Let us suppose that the state $\rho(\theta)$ depends on three unknown parameters. Then we can parameterize it by $\rho(\theta) = \frac{1}{2}(I + \sum_i \theta_i \sigma_i)$ where $I$ is the identity matrix, $\sigma_i$ are the Pauli matrices, and the three parameters $\theta_i$ obey $\|\theta\|^2 = \sum_i \theta_i^2 \leq 1$. We now introduce the tensor

$$H_{ij}(\theta) = \delta_{ij} + \frac{\theta_i \theta_j}{1 - \|\theta\|^2} \quad (8)$$

which generalizes tensor (6) to the case of mixed states. Then, upon restricting oneself to separable measurements, we will show that the MQE matrix $V^N$ must satisfy (exactly for unbiased estimators, and otherwise asymptotically)

$$\text{tr} \ H(\theta)^{-1} V^N(\theta)^{-1} \leq N \quad . \quad (9)$$
As an application of these results, the minimum of the cost function (3) in the case of spin-1/2 particles (for mixed states restricting oneself to separable measurement) is

\[
\min_{\theta} E_{\theta}(f(\hat{\theta}, \theta)) = f_0(\theta) + \left(\frac{\text{tr} \sqrt{H(\theta)^{-1/2}C(\theta)H(\theta)^{-1/2}}} N\right)^2 + o(1/N) \quad (10)
\]

which is obtained simply by minimizing Eq. (3) subject to the constraints (7) or (9).

We can compare Eq. (10) with the exact results which are known in the case of covariant measurements on pure states of spin-1/2 particles [9, 15]. In this problem one is given \(N\) spin-1/2 particles polarized along the direction \(\Omega\). \(\Omega\) is uniformly distributed on the sphere. One wants to devise a measurement and estimation strategy that minimize the mean value of the cost function \(\cos^2 \omega/2\), where \(\omega\) is the angle between the estimated direction \(\hat{\Omega}\) and the true direction \(\Omega\). Expanding the cost function to second order in \(\omega\) (to obtain the quadratic cost matrix \(C\)) and averaging Eq. (10) over the sphere, one finds

\[
E(\cos^2 \omega/2) \geq 1 - \frac{1}{N} + o\left(\frac{1}{N}\right) \quad (11)
\]

which in the limit for large \(N\) coincides with the results (exact for all \(N\)) of Refs. [9, 15]. If the directions \(\Omega\) are not uniformly distributed, then Refs. [9, 15] do not apply, but Eq. (10) stays valid. However, we cannot compare our results with the recent analysis of covariant measurements on mixed states [20] because we suppose separability of the measurement, whereas Ref. [20] does not.

Equations (7) and (9) have a simple generalization to the case of particles belonging to higher-dimensional Hilbert spaces. But in these cases these bounds are no longer sharp.

In order to appreciate the above results, we must recall some results from classical statistical inference. This is the subject of Sec. 2.

2. Classical Cramér-Rao Bound

Consider a random variable \(X\) with probability density \(p(x, \theta)\). The connection with the quantum problem is that we can view \(p(x, \theta)\) as the probability density that a quantum measurement on the system yields an outcome \(x\) given that the state was \(\rho(\theta)\). We take a random sample of size \(N\) from the distribution, and use it to estimate the value of each parameter \(\theta_i\). Call \(\hat{\theta}_i^N\)
the estimated value. The following results about the MQE matrix of the estimator are well known.

(1) Suppose that the estimator is unbiased, that is, 
\[ E_\theta(\hat{\theta}^N - \theta) = 0, \]
where \( E_\theta \) is the expectation value at fixed \( \theta \), i.e., the integral \( \int dx p(x|\theta) \).

Define its MQE matrix \( V_N(\theta) \) by
\[ V_{ij}^N(\theta) = E_\theta((\hat{\theta}_i^N - \theta_i)(\hat{\theta}_j^N - \theta_j)). \]  

Furthermore define the Fisher information matrix \( I(\theta) \) by
\[ I_{ij}(\theta) = E_\theta(\frac{\partial}{\partial \theta_i} \ln p(X|\theta) \frac{\partial}{\partial \theta_j} \ln p(X|\theta)) = \int dx \frac{\partial}{\partial \theta_i} p(x|\theta) \frac{\partial}{\partial \theta_j} p(x|\theta) \]
\[ p(x|\theta), \]  

Then, for any \( N \), the following inequalities, known as the Cramér-Rao inequalities, hold [4,8]:
\[ V^N(\theta) \geq I(\theta)^{-1}/N \]  
or, equivalently
\[ V^N(\theta)^{-1} \leq NI(\theta), \]
the inequality meaning that the difference of the two sides is a non-negative matrix.

(2) The hypothesis of unbiased estimators is very restrictive since most estimators will be biased. Happily it is possible to relax this condition. Here are just two of the many results available:

(a) First of all, if one is interested in averaging the mean cost over possible values of \( \theta \) with respect to a given prior distribution \( \lambda(\theta) \), then there is a Bayesian version of the Cramér-Rao inequality, the van Trees inequality [7,19]. In the multivariate case, upon giving oneself a quadratic cost function determined by a matrix \( C(\theta) \), one can derive the inequality
\[ \int d\theta \lambda(\theta) \text{tr} \ C(\theta)V^N(\theta) \geq \frac{\int d\theta \lambda(\theta) \text{tr} \ C(\theta)I^{-1}(\theta)}{N} - \frac{\alpha}{N^2} \]  
where \( \alpha \) is a positive number that depends on \( C(\theta), I(\theta), \) and \( \lambda(\theta) \), but is independent of \( N \).

(b) The second approach makes no reference to any prior distribution for \( \theta \), but only holds in the limit \( N \) tending to infinity and lays a mild restriction on the estimators considered. Specifically, if the probability distribution of \( \sqrt{N}(\hat{\theta}^N - \theta) \) converges uniformly in \( \theta \) towards a distribution depending continuously on \( \theta \), say of a random
vector $Z$, then the limiting scaled MQE matrix $W(\theta)$ defined by $W_{ij}(\theta) = E_\theta(Z_i Z_j)$ obeys $W \geq I^{-1}$.

(3) Furthermore in the limit of arbitrarily large samples one can attain the Cramér-Rao bound. This is proven by explicitly constructing an estimator that attains the bound in the extended senses (2a) (apart from the $1/N^2$ term) or (2b) just indicated: the maximum likelihood estimator (MLE).

Modern statistical theory contains many other results having the same flavour as point (2) above, namely, that the Cramér-Rao bound holds in an approximate sense for large $N$, without the restriction to biased estimators. Result (2a) applies to a larger class of estimators than (2b), but only gives a result on the average behaviour over different values of $\theta$. On the other hand combining results (3) and (2b) tells us that the maximum likelihood estimator is for large $N$ an optimal estimator for each value of $\theta$ separately. The reason why in (2b) additional regularity is demanded is because of the phenomenon of super-efficiency (see Ref. [18] for a recent discussion), whereby an estimator can have mean quadratic error of smaller order than $1/N$ at isolated points. Modern statistical theory (see again Refs. [18] or [11]) has concentrated on the more difficult problem of obtaining non-Bayesian results (i.e., pointwise rather than average) making much use of the technical tool of “local asymptotic normality.” A major challenge in the quantum case is to obtain a result of type (2b) when this technique is definitely not available.

3. Quantum Cramér-Rao Bound

In this paper we show that similar results to (1), (2a), (2b), and (3) can be obtained when one must estimate the state of an unknown quantum system $\rho(\theta)$ of which one possesses $N$ copies. This problem is most simply addressed, following Ref. [3], by decomposing it into a first (quantum) step in which one carries out a measurement on $\rho^N = \rho \otimes \cdots \otimes \rho$ and a second (classical) step in which one uses the result of the measurement to estimate the value of the parameters $\theta$.

The most general way to describe the measurement is by a positive operator-valued measurement (POVM) $M = (M_\xi)$ whose elements satisfy $M_\xi \geq 0$, $\sum_\xi M_\xi = I$. (For simplicity we take the outcomes of the POVM to be discrete. The generalization to an arbitrary outcome space is just a question of translating into measure-theoretic language.)
Quantum mechanics tells us the probability to obtain outcome $\xi$ given state $\rho$:

$$p(\xi|\theta) = \text{Tr} \rho^N(\theta) M_\xi.$$  \hfill (17)

From the outcome $\xi$ of the measurement one must guess what are the values of the $p$ parameters $\theta_i$. Call $\hat{\theta}^N$ the estimated value of the parameter vector. We want to obtain bounds on the MQE matrix $V^N(\theta)$ of the estimator $\hat{\theta}^N$ when the true parameter value is $\theta$, thus $V^{N}_{ij}(\theta) = \text{E}_\theta(\hat{\theta}^N_i - \theta_i)(\hat{\theta}^N_j - \theta_j)$. To proceed we temporarily make the simplifying assumption that the estimators are unbiased, $\text{E}_\theta \hat{\theta}^N = \theta$. Then we can apply the classical Cramér-Rao inequality to the probability distribution $p(\xi|\theta)$ to obtain

$$V^N \geq I^N(M, \theta)^{-1}$$  \hfill (18)

or

$$(V^N)^{-1} \leq I^N(M, \theta)$$  \hfill (19)

where the Fisher information matrix $I^N$ for the measurement $M$ is defined by

$$I^N_{ij}(M, \theta) = \sum_{\xi} \frac{\partial_i p(\xi|\theta) \partial_j p(\xi|\theta)}{p(\xi|\theta)} = \sum_{\xi} \frac{\text{Tr}(\rho^N_{ij} M_\xi) \text{Tr}(\rho^N_{ji} M_\xi)}{\text{Tr}(M_\xi \rho^N)}$$  \hfill (20)

with $\rho^N_{ij} = \partial_{\theta_i} \rho^N$.

These expressions suggest the following questions.

1. is there a simple bound for the MQE $V^N$ of unbiased estimators $\hat{\theta}^N$, or equivalently for the Fisher information $I^N(M, \theta)$?
2. is the bound also valid for sufficiently well behaved but possibly biased estimators—at least in the limit of large $N$?
3. can this bound be attained—at least in the limit of a large number of copies $N$?

Most of the work on this subject has been devoted to answering the question (1). We now recall what is known about these questions.

Suppose first the parameter $\theta$ is one-dimensional, $p = 1$. The symmetric logarithmic derivative (SLD) $\lambda_\theta$ of $\rho$ is the Hermitian matrix defined implicitly by

$$\rho_{,\theta} = \frac{\lambda_\theta \rho + \rho \lambda_\theta}{2}.$$  \hfill (21)
In a basis where $\rho$ is diagonal, $\rho = \sum_k p_k |k\rangle \langle k|$, this can be inverted to yield
\[
(\lambda_\theta)_{kl} = (\rho, \theta)_{kl} \frac{2}{p_k + p_l}. \tag{22}
\]

Then we have the bound
\[
I^N_{\theta\theta}(M, \theta) \leq N \text{Tr} \rho \lambda_\theta \lambda_\theta. \tag{23}
\]

Furthermore it was suggested in Ref. [1] how to adapt the classical MLE so as to attain, in the limit of large $N$, the bound (23).

In the multiparameter case the bound based on the SLD can be generalized in a natural way. Define the SLD along direction $\theta_i$ by
\[
\rho^{\theta,i} = \frac{\lambda_i \rho + \rho \lambda_i}{2}, \tag{24}
\]
and Helstrom’s quantum information matrix $H$ by
\[
H_{ij} = \text{Tr} \rho \frac{\lambda_i \lambda_j + \lambda_j \lambda_i}{2}. \tag{25}
\]

(This is the same matrix that was introduced for spin-1/2 particles for a particular choice of parameters in (6) and (8)). Then one can prove the bound [8],
\[
I^N(M, \theta) \leq N H(\theta). \tag{26}
\]

(This can be deduced directly from Eq. (23) as proven in [3]. Indeed since Eq. (23) holds for each path in parameter space, it implies the matrix equation (26)).

However, this bound is in general not achievable. Another bound has been proposed based on an asymmetric logarithmic derivative (ALD) [24] which in some cases is better than Eq. (26). Holevo [9] has proposed yet another bound that is stronger than both the SLD and the ALD bounds, but this bound is not explicit: it requires a further minimization. As far as we know no general achievable bound is known in the multiparameter case.

The difficulty in obtaining a simple bound in the multiparameter case is that there are many inequivalent ways in which one can minimize the MQE matrix $V^N_{ij}$. That is, in order to build a good estimator one must make a choice of what one wants to estimate, and according to this choice the measurement strategy followed will be different. Hence a bound in the form of a matrix inequality like Eq. (26) cannot be expected to be tight.
4. Results

In this paper we obtain answers to the three questions raised above in the multiparameter case. Our results are summarized in this section.

We first discuss point 1), that is bounds on the Fisher information. We shall show the following:

**Theorem 1:** When \( \rho(\theta) = |\psi(\theta)\rangle\langle\psi(\theta)| \) is a pure state, then the Fisher information \( I^N(M, \theta) \) defined in Ref. (20) must satisfy the relation

\[
\text{tr} \, H^{-1}(\theta) I^N(M, \theta) \leq (d - 1)N \tag{27}
\]

where \( H^{-1} \) is the inverse of the quantum information matrix defined in Eq. (25), and \( d \) is the dimension of the Hilbert space to which \( \rho(\theta) \) belongs. Note that this inequality (27) is invariant under change of parameterization \( \theta \rightarrow \theta'(\theta) \).

This result immediately gives an inequality for the mean quadratic error matrix of unbiased estimators \( \hat{\theta}^N \) by invoking the classical Cramér-Rao inequality in order to replace \( I^N(M, \theta) \) by the inverse of the MQE \( V^N(\theta) \):

\[
\text{tr} \, H^{-1}(\theta)(NV^N(\theta))^{-1} \leq d - 1 \tag{28}
\]

**Theorem 2:** When \( \rho(\theta) \) is a mixed state, and if the measurement \( M \) consists of separate measurements on each particle, then the Fisher information also satisfies Eq. (27). Hence for separable measurements on a mixed state, the MQE matrix of an unbiased estimator satisfies Eq. (28).

**Theorem 3:** In the case of mixed states, it is in general possible to devise a collective measurement for which the Fisher information does not satisfy the inequality (27).

The second part of the paper consists of proving that the constraint (28) also holds for biased estimators under suitable additional conditions. We give two forms of this generalized form of Eq. (28) corresponding to the two forms (2a) and (2b) of the generalized classical Cramér-Rao inequality.

Consider \( N \) copies of a state \( \rho(\theta) \). If \( \rho \) is pure we can make either collective or separable measurements. If \( \rho \) is mixed we restrict ourselves to separable measurements [since Theorem 3 shows that in this case collective measurements can beat Eq. (27)]. Based on the outcome of the measurement we estimate the value of the parameter vector \( \theta \). Call \( \hat{\theta} \) the estimator, and denote by \( V^N = V^N(\theta) \) its MQE matrix when the true value of the parameter is \( \theta \).
We shall prove the following generalization of result of type 2b) concerning the behaviour of the mean quadratic error matrix as \( N \) tends to infinity:

**Theorem 4:** Suppose that the scaled MQE \( NV^N(\theta) \) has the limit \( W(\theta) \) as \( N \rightarrow \infty \). Suppose that the convergence is uniform in \( \theta \), and that \( W \) is continuous at the point \( \theta = \theta^0 \). Furthermore we suppose that \( H \) and its derivatives are bounded in a neighbourhood of this point. Then we shall prove in section 6 that \( W(\theta^0) \) must satisfy

\[
\text{tr} H^{-1}(\theta^0)W^{-1}(\theta^0) \leq (d - 1).
\] (29)

This result gives a bound on the mean value of a quadratic cost function \( C \) as \( N \) tends to infinity. Indeed, using a Lagrange multiplier to impose the condition (29), the minimum cost is readily found to be

\[
\lim_{N \rightarrow \infty} N \text{tr} C(\theta^0)V^N(\theta^0) \geq \frac{1}{d - 1} \left( \text{tr} \sqrt{H^{-\frac{1}{2}}(\theta^0)C(\theta^0)H^{-\frac{1}{2}}(\theta^0)} \right)^2.
\] (30)

In terms of a cost function, it is also possible to prove a Bayesian version of the Cramér-Rao inequality which is the analogue of the classical result (2a).

**Theorem 5:** Suppose that one is given a quadratic cost function \( C(\theta) \) and a prior distribution \( \lambda(\theta) \) for the parameters \( \theta \). If \( C, \lambda, \) and \( H \) are sufficiently smooth functions of \( \theta \) (the continuity of the first derivatives is sufficient), while \( \lambda \) is zero outside a compact region with smooth boundary, then

\[
\int d\theta \lambda(\theta) \text{tr} C(\theta)V^N(\theta) \\
\geq \frac{1}{(d - 1)N} \int d\theta \lambda(\theta) \text{tr} \left( \sqrt{H^{-\frac{1}{2}}(\theta)C(\theta)H^{-\frac{1}{2}}(\theta)} \right)^2 - \frac{\alpha N^2}{N^2} \] (31)

where \( \alpha \) is a constant independent of \( N \) but which depends on \( C, \lambda \) and \( H \).

Theorems 1, 2, 4, and 5 put bounds on the MQE matrix of an estimator of an unknown state \( \rho(\theta) \) (for mixed states, under the restriction that the measurement is separable). The third part of this paper is devoted to showing that in the case of spin-1/2 systems (\( d = 2 \)) these bounds can be attained. We first show that at any point \( \theta^0 \) we can attain equality in Eq. (27).

**Theorem 6:** Suppose one has \( N \) spin-1/2 particles in an unknown (possibly mixed) state \( \rho(\theta) \). Fix any point \( \theta^0 \). Give yourself a matrix \( G^0 \) satisfying
tr $H^{-1}(\theta^0)G^0 \leq 1$. We call $G^0$ the target scaled information matrix. Then there exists a measurement $M^{\theta^0}$ (depending on the choice of $\theta^0$) acting on each spin separately such that $I^N(M^{\theta^0}, \theta^0) = NG^0$. This measurement is described in detail in Sec. 7.1.

For large $N$ we can also approximately attain equality at all points $\theta$ simultaneously.

**Theorem 7:** Suppose one has $N$ spin-1/2 particles in an unknown pure state $|\psi(\theta)\rangle$, or suppose that one has $N$ spin-1/2 particles in an unknown mixed state $\rho(\theta)$. In the latter case we also require that the state never be pure, i.e., $\text{Tr} \rho(\theta)^2 < 1$ for all $\theta$.

Give oneself a smooth positive matrix function $G(\theta)$ satisfying $\text{tr} H^{-1}(\theta)G(\theta) \leq 1$ for all $\theta$, the target scaled information for each possible value of $\theta$. Define the corresponding target scaled MQE matrix $W(\theta) = G(\theta)^{-1}$. Suppose that $W(\theta)$ is nonsingular [i.e., $G(\theta)$ never has a zero eigenvalue]. Then there exists a measurement $M$ acting on each spin separately, and a corresponding estimator $\hat{\theta}$, whose MQE matrix $V^N(\theta)$ satisfies

$$V^N(\theta) = \frac{W(\theta)}{N} + o(1/N)$$

for all values of $\theta$ simultaneously. For this estimation strategy $\sqrt{N}(\hat{\theta} - \theta)$ converges in distribution towards $N(0, W)$, the normal distribution with mean zero and covariance $W$. The measurement $M$ and estimation strategy is described in detail in Sec. 7.2.

It is interesting to note that the measurement strategy which satisfies Eq. (32) is an adaptive one. That is, one first carries out a measurement on a small fraction of the particles. This gives a preliminary estimate of the quantum state which allows a fine tuning of the measurements that are carried out on the remaining particles. This is to be contrasted with previously proposed state estimation strategies in the case of finite-dimensional Hilbert spaces [14, 22] in which the same measurement is carried out on all the particles. The necessity of an adaptive measurement strategy if one wants to minimise the MQE was pointed out in Ref. [1].

When the unknown state belongs to a Hilbert space of dimension $d > 2$, then bound (27) cannot be attained in general. Indeed, we shall show in Sec. 5.6 that for $d > 2$, neither Eq. (26) nor (27) implies the other.
5. New Quantum Cramér-Rao Inequality

In this section we prove Theorems 1, 2, and 3. That is, we prove Eq. (27) for general measurements in the case of pure states and for separate measurements on each particle in the case of mixed states.

5.1. Preliminary results

The first step in proving Eq. (27) is to show that one can restrict oneself to POVM’s whose elements are proportional to one-dimensional projectors. Indeed, any POVM can always be refined to yield a POVM whose elements are proportional to one dimensional projectors. We call such a measurement exhaustive. This yields a refined probability distribution $(p(\xi, \theta))$. It is well known that under such refining of the probability distribution, the Fisher information can only increase [5].

The second step in proving Eq. (27) consists of increasing the number of parameters. Suppose that $\rho(\theta)$ depends on $p$ parameters $\theta_i$, $i = 1, \ldots, p$. If $\rho = |\psi(\theta))\langle\psi(\theta)|$ is a pure state, then $p \leq 2d - 2$ (since $|\psi(\theta)|$ is normalized and defined up to a phase). If $\rho$ is a mixed state, then Hermiticity and the condition $\text{Tr} \rho = 1$ impose that $p \leq d^2 - 1$. Suppose that $p < \nu$ is less then the maximum number of possible parameters ($\nu = 2d - 2$ or $\nu = d^2 - 1$ according to whether the state is pure or mixed). Then one can always increase the number of parameters up to the maximum. Indeed let us suppose that to the $p$ parameters, one adds independent parameters $\theta_{i'}$, $i' = p + 1, \ldots, \nu$. We can now consider the quantum information matrix $\tilde{H}$, and Fisher information matrix $\tilde{I}$, for the completed set of parameters. We shall show below that

$$\text{tr} H^{-1}(\theta) I^N(M, \theta) \leq \text{tr} \tilde{H}^{-1}(\theta) \tilde{I}^N(M, \theta).$$

(33)

Therefore it will be sufficient to prove Eq. (27) in the case when there are $\nu$ parameters.

To prove Eq. (33), fix a particular point $\theta^0$. At this point we have the derivative $\rho_i$ and SLD $\lambda_i$ of $\rho$ for $i = 1, \ldots, p$. Introduce a set of Hermitian matrices $\lambda_i'$ with $\text{Tr} \rho(\theta^0)\lambda_i' = 0$, for $i' = p + 1, \ldots, \nu$, such that

$$\text{Tr} \rho(\theta^0)\frac{\lambda_i'\lambda_i + \lambda_i\lambda_i'}{2} = 0 \quad , \quad i = 1, \ldots, p \quad , \quad i' = p + 1, \ldots, \nu. \quad (34)$$

This is always possible because we can view (34) as a scalar product between $\lambda_i$ and $\lambda_i'$ and a Gram-Schmidt orthogonalization procedure will then yield the matrices $\lambda_i'$. Now define matrices $\rho_{i'}$ by $\rho_{i'} = (\rho(\theta^0)\lambda_i' + \lambda_i\rho(\theta^0))/2$ and define additional parameters $\theta_{i'}$ satisfying, at $\theta^0$: $\partial_{\theta_{i'}} \rho = \rho_{i'}$. The point
of this construction is that because of Eq. (34), the quantum information
matrix $\tilde{H}$ is block diagonal with the first block equal to $H$. Let $\tilde{I}(M)$ be the
Fisher information matrix for the enlarged set of parameters (but the same
measurement). Then $\text{tr} \tilde{H}^{-1} \tilde{I}(M) = \text{tr}(\tilde{H}^{-1})_{11} \tilde{I}_{11}(M) + \text{tr}(\tilde{H}^{-1})_{22} \tilde{I}_{22}(M)$
where the indices 11 and 22 denote the blocks of these matrices correspond-
ing to the original and new parameters. But both terms are non-negative
since all matrices involved are non-negative, and $(\tilde{H}^{-1})_{11} = H^{-1}$, so we ob-
tain Eq. (33) at $\theta_0$ and for the particular parameters just introduced. But
since the right-hand side of Eq. (33) is invariant under reparameterization,
it is true for any parameterization, and at any $\theta$.

5.2. Pure states

To proceed we shall consider a POVM whose elements are proportional
to one dimensional projectors, and calculate explicitly the left-hand side
of Eq. (27) in the case where the number of parameters is the maximum
$p = 2d - 2$ in a basis where $H$ is diagonal.

We fix a point $\theta^0$. At this point we chose a basis such that
\begin{equation}
\rho(\theta^0) = |1\rangle\langle 1| . \tag{35}
\end{equation}

Hence the density matrix of the $N$ copies is
\begin{equation}
\rho^N = |1\rangle\langle 1| \otimes \ldots \otimes |1\rangle\langle 1| . \tag{36}
\end{equation}

Consider now the $2d - 2$ Hermitian operators
\begin{equation}
\rho,_{k+} = |1\rangle \langle k| + |k\rangle \langle 1|, \quad \rho,_{k-} = i |1\rangle \langle k| - i |k\rangle \langle 1|, \quad 1 < k \leq d . \tag{37}
\end{equation}

We choose a parameterization such that in the vicinity of $\theta^0$, it has the
form $\rho = \rho(\theta^0) + \sum_{k,\pm} (\theta_{k\pm} - \theta^0_{k\pm}) \rho,_{k\pm}$ with the unknown parameters $\theta_{k\pm}$,
$k = 2, \ldots, d$. With this parameterization the derivatives of $\rho^N$ are
\begin{equation}
\rho^N,_{k\pm} = \rho,_{k\pm} \otimes \rho \otimes \cdots \otimes \rho \otimes \cdots \otimes \rho,_{k\pm} . \tag{38}
\end{equation}

One then calculates the SLD of $\rho$, and hence the quantum information
matrix $H$. One verifies that in this basis $H$ is diagonal:
\begin{equation}
H_{k\pm, k'\pm'} = 4 \delta_{kk'} \delta_{\pm\pm'} . \tag{39}
\end{equation}

Consider any POVM whose elements are proportional to one dimen-
sional projectors
\begin{equation}
M_\xi = |\psi_\xi \rangle \langle \psi_\xi | , \quad |\psi_\xi \rangle = \sum_{k_1=1}^d \ldots \sum_{k_N=1}^d a_{\xi k_1 \ldots k_N} |k_1 \ldots k_N \rangle . \tag{40}
\end{equation}
The completeness relation \( \sum \xi M_\xi = I \) takes the form
\[
\sum_\xi a_{\xi k_1...k_N}^* a_{\xi' k'_1...k'_N} = \delta_{k_1 k'_1} \ldots \delta_{k_N k'_N}.
\] (41)

To proceed we need the formulas
\[
\text{Tr } \rho(\theta^0) M_\xi = |a_{\xi 1...1}|^2 \tag{42}
\]
and
\[
\text{Tr } \rho(\theta^0),_{k+} M_\xi = \sum_{p=1}^{N} (a_{\xi 1...1}^* a_{\xi 1...k_p=k...1} + a_{\xi 1...k_p=k...1}^* a_{\xi 1...1}), \tag{43}
\]
and similarly for \( \text{Tr } \rho(\theta^0),_{k-} M_\xi \). Thus we obtain
\[
(\text{Tr } \rho(\theta^0),_{k+} M_\xi)^2 + (\text{Tr } \rho(\theta^0),_{k-} M_\xi)^2 = \sum_{p=1}^{N} 4 |a_{\xi 1...1}|^2 |a_{\xi 1...k_p=k...1}|^2. \tag{44}
\]

Putting everything together yields
\[
\text{tr } H^{-1} I(M) = \sum_\xi \frac{1}{\text{Tr } \rho(\theta^0) M_\xi} \frac{1}{4} \sum_{k=2}^{d} \sum_{\pm} (\text{Tr } \rho(\theta^0),_{k+} M_\xi)^2 + (\text{Tr } \rho(\theta^0),_{k-} M_\xi)^2
\]
\[
= \sum_{k=2}^{d} \sum_{p=1}^{N} \sum_\xi |a_{\xi 1...k_p=k...1}|^2 = N(d - 1), \tag{45}
\]
which proves that equality holds in Eq. (27) for arbitrary exhaustive measurements in the case of pure states.

5.3. One mixed state

Deriving Eq. (27) for mixed states is more complicated than for pure states, and we shall proceed in two steps. First we shall consider the case of one mixed state \((N = 1)\), and show that equality in Eq. (27) holds in this case for arbitrary exhaustive measurements. Then we shall consider the case of an arbitrary number \(N\) of mixed states.

We first diagonalize \(\rho\) at a point \(\theta^0\): \(\rho(\theta^0) = \sum_{k=1}^{d} p_k |k\rangle \langle k|\). We now introduce the following complete set of Hermitian traceless matrices:
\[
\rho,_{kl+} = |k\rangle \langle l| + |l\rangle \langle k|, \quad \rho,_{kl-} = i |k\rangle \langle l| - i |l\rangle \langle k|, \quad k < l,
\]
\[
\rho,_{m} = \sum_{k=1}^{d} c_{mk} |k\rangle \langle k|, \quad m = 1, \ldots, d - 1, \tag{46}
\]
where the coefficients \( c_{mk} \) obey
\[
\sum_k c_{mk} = 0, \quad \sum_k \frac{1}{p_k} c_{m'k} c_{mk} = \delta_{m'm}.
\] (47)

Let us denote the matrices \( \rho_{kl\pm} \) and \( \rho_m \) collectively as \( \rho_i \). [They constitute a set of generators of \( su(d) \)].

We choose a parameterization such that in the vicinity of \( \theta^0 \), it has the form \( \rho = \rho(\theta^0) + \sum_i (\theta_i - \theta^0_i) \rho_i \). One then calculates the SLD of \( \rho \), and from this the quantum information matrix \( H \). One verifies that in this basis \( H \) is diagonal:
\[
H_{kl\pm,k'l'\pm} = \frac{4}{p_k + p_l} \delta_{kk'} \delta_{ll'} \delta_{\pm\pm'}, \quad H_{kl\pm,m} = 0, \quad H_{m,m'} = \delta_{m'm}.\] (48)

Consider any POVM whose elements are proportional to one dimensional projectors
\[
M_\xi = |\psi_\xi \rangle \langle \psi_\xi |, \quad |\psi_\xi \rangle = \sum_k a_{\xi k} |k\rangle.
\] (49)

The left-hand side of Eq. (27) can now be written as
\[
\text{tr} H^{-1} I(M) = \sum_\xi \frac{1}{\langle \psi_\xi | \rho | \psi_\xi \rangle} \left( \sum_{k<l} \frac{p_k + p_l}{4} \langle \psi_\xi | \rho_{kl\pm} | \psi_\xi \rangle^2 + \sum_m \langle \psi_\xi | \rho_m | \psi_\xi \rangle^2 \right).\] (50)

Using the following expressions
\[
\langle \psi_\xi | \rho_m | \psi_\xi \rangle = \sum_k |a_{\xi k}|^2 c_{mk}, \quad \langle \psi_\xi | \rho_{kl+} | \psi_\xi \rangle^2 + \langle \psi_\xi | \rho_{kl-} | \psi_\xi \rangle^2 = 4 |a_{\xi k}|^2 |a_{\xi l}|^2,
\]
on one obtains
\[
\text{tr} H^{-1} I(M) = \sum_\xi \frac{1}{\langle \psi_\xi | \rho | \psi_\xi \rangle} \left( \sum_{k<l} (p_k + p_l) |a_{\xi k}|^2 |a_{\xi l}|^2 + \sum_m \left( \sum_k |a_{\xi k}|^2 c_{mk} \right)^2 \right)
= \sum_\xi \frac{1}{\langle \psi_\xi | \rho | \psi_\xi \rangle} \left( \sum_{k\neq l} p_k |a_{\xi k}|^2 |a_{\xi l}|^2 + \sum_k \sum_l |a_{\xi k}|^2 |a_{\xi l}|^2 \sum_m c_{mk} c_{ml} \right).
\] (51)

We now use the relation
\[
\sum_m c_{mk} c_{ml} = \delta_{kl} p_k - p_k p_l,
\] (52)
which is derived from (47) as follows: define \( v_{mk} = c_{mk} / \sqrt{p_k} \) (\( m = 1, \ldots, d-1 \)) and \( v_{dk} = \sqrt{p_k} \). Then (47) can be rewritten as \( \sum_k v_{mk} v_{m'k} = \delta_{mm'} \). The vectors \( v_{mk} \) therefore are a complete orthonormal basis of \( \mathbb{R}^d \), hence they obey \( \sum_m v_{mk} v_{mk} = \delta_{kk'} \). Reexpressing in terms of \( c_{mk} \) yields Eq. (52). Inserting it in Eq. (51) we obtain

\[
\text{tr} H^{-1} I(M) = \sum_\xi \frac{1}{\langle \psi_\xi | \rho | \psi_\xi \rangle} \left( \sum_k \sum_l p_k (1 - p_l) |a_{\xi k}|^2 |a_{\xi l}|^2 \right)
= \sum_k (1 - p_k) \sum_\xi |a_{\xi k}|^2 = \sum_\xi \text{Tr}(I - \rho) M_\xi = d - 1,
\]

as announced.

Note that this has demonstrated that equality holds in (27) whenever \( N = 1 \), \( p = d^2 - 1 \), and the POVM is exhaustive. It follows from the classical properties of the Fisher information that equality also holds for arbitrary \( N \) whenever the POVM can be considered as a sequence of \( N \) separate exhaustive measurements on each copy of the system. It also holds if the \( n \)th measurement is chosen at random depending on the outcomes of the previous measurements.

5.4. Separable measurements on \( N \) mixed states

We shall now prove that if we possess \( N \) identical mixed states of spin-1/2 particles, and carry out separable measurements, then

\[
\text{tr} H^{-1} I(M) \leq N(d - 1).
\]

We recall that a separable measurement is one that can be carried out sequentially on separate particles, where the measurement on one particle at any stage (and indeed which particle to measure: one is allowed to measure particles several times) can depend arbitrarily on the outcomes so far; see Ref. [16] for a discussion. It is therefore more general than the case considered at the end of the previous subsection where the measurement on the \( n \)th particle could only depend on the measurements carried out on the \( n - 1 \) previous particles.

If a POVM is separable, then its elements \( M_\xi \) can be decomposed into a sum of terms proportional to projectors onto unentangled states

\[
M_\xi = \sum_i |\psi_{\xi i} \rangle \langle \psi_{\xi i}|, \quad |\psi_{\xi i} \rangle = |\psi_{\xi i}^1 \rangle \otimes \cdots \otimes |\psi_{\xi i}^N \rangle.
\]

We call measurements having such a representation nonentangled. (Note that there exist nonentangled POVM’s which are not separable [2].).
By refining a separable measurement (which increases the Fisher information), one can restrict oneself to measurements whose POVM elements are proportional to projectors onto product states:

\[ M_\xi = |\psi_\xi\rangle\langle\psi_\xi| = |\psi_\xi^1\rangle\langle\psi_\xi^1| \otimes \cdots \otimes |\psi_\xi^N\rangle\langle\psi_\xi^N|. \]  

(56)

We now evaluate the left-hand side of Eq. (54) for measurements of form (56). First recall that the \( \rho \) and the derivatives of \( \rho \) are given by

\[ \rho^N = \rho \otimes \cdots \otimes \rho = \sum_{k_1=1}^{d} \cdots \sum_{k_N=1}^{d} p_{k_1} \cdots p_{k_N} |k_1 \cdots k_N\rangle \langle k_1 \cdots k_N|, \]  

(57)

and the derivatives of \( \rho^N \) have the form

\[ \rho_{,i}^N = \rho_{,i} \otimes \cdots \otimes \rho + \cdots + \rho \otimes \cdots \otimes \rho_{,i} = \sum_{p=1}^{N} \rho \otimes \cdots \otimes \rho_{,i} \cdots \otimes \rho, \]  

(58)

where in the second rewriting it is understood that \( \rho_{,i} \) is at the \( p \)th position in the product.

Using the product form of measurement (56), one finds that

\[
\langle \psi_\xi | \rho^N | \psi_\xi \rangle = \langle \psi_\xi^1 | \rho | \psi_\xi^1 \rangle \cdots \langle \psi_\xi^N | \rho | \psi_\xi^N \rangle,
\]

\[
\langle \psi_\xi | \rho_{,i}^N | \psi_\xi \rangle = \sum_{p=1}^{N} \langle \psi_\xi^1 | \rho | \psi_\xi^1 \rangle \cdots \langle \psi_\xi^P | \rho_{,i} | \psi_\xi^P \rangle \cdots \langle \psi_\xi^N | \rho | \psi_\xi^N \rangle.
\]  

(59)

Inserting these expressions into the Fisher information matrix, one finds

\[
I_{ij}(M) = \sum_{\xi} \frac{\langle \psi_\xi | \rho_{,i}^N | \psi_\xi \rangle \langle \psi_\xi | \rho_{,j}^N | \psi_\xi \rangle}{\langle \psi_\xi | \rho^N | \psi_\xi \rangle}
= \sum_{\xi} \sum_{p \neq p'} \langle \psi_\xi^1 | \rho | \psi_\xi^1 \rangle \cdots \langle \psi_\xi^P | \rho_{,i} | \psi_\xi^P \rangle \cdots \langle \psi_\xi^N | \rho | \psi_\xi^N \rangle
+ \sum_{\xi} \sum_{p} \langle \psi_\xi^1 | \rho | \psi_\xi^1 \rangle \cdots \frac{\langle \psi_\xi^P | \rho_{,i} | \psi_\xi^P \rangle \langle \psi_\xi^P | \rho_{,j} | \psi_\xi^P \rangle}{\langle \psi_\xi^P | \rho | \psi_\xi^P \rangle} \cdots \langle \psi_\xi^N | \rho | \psi_\xi^N \rangle
= \sum_{\xi} \sum_{p} \langle \psi_\xi^1 | \rho | \psi_\xi^1 \rangle \cdots \frac{\langle \psi_\xi^P | \rho_{,i} | \psi_\xi^P \rangle \langle \psi_\xi^P | \rho_{,j} | \psi_\xi^P \rangle}{\langle \psi_\xi^P | \rho | \psi_\xi^P \rangle} \cdots \langle \psi_\xi^N | \rho | \psi_\xi^N \rangle,
\]  

(60)

where we have used the fact that the first term in the second equality vanishes. Indeed it is equal to

\[
\sum_{\xi} \sum_{p \neq p'} \langle \psi_\xi | \rho \otimes \cdots \otimes \rho_{,i} \otimes \cdots \otimes \rho_{,j} \otimes \cdots \otimes \rho | \psi_\xi \rangle.
\]  

(61)
The sum over $\xi$ can be carried out in Eq. (61) to yield the identity matrix and the resulting trace vanishes because $\text{Tr} \rho \otimes \ldots \otimes \rho = 0$.

We now insert Eq. (60) into $\text{tr} H^{-1}(M)$. All the operations from Eqs. (50)-(53) can be carried out exactly as in the previous subsection, and one arrives at the expression

$$\text{tr} H^{-1}(M) = \sum_p \sum_\xi \langle \psi_\xi | \rho \otimes \ldots \otimes (I - \rho) \otimes \ldots \otimes \rho | \psi_\xi \rangle = N(d - 1)$$

which is the sought for relation.

5.5. Inequality for more then one mixed state

We now provide a counterexample showing that if one carries out a collective measurement on $N > 1$ mixed states one can violate Eq. (27). We take $N = 2$, and suppose the unknown states belong to a 2 dimensional Hilbert space. $\rho(\theta) = \frac{1}{2}(I + \sum_i \theta_i \sigma_i)$. We take as reference point $\theta_i = 0$ corresponding to $\rho = \frac{1}{2} I$. At this point $H_{ij}(\theta_i = 0) = \delta_{ij}$.

We consider as measurement on the two copies the following POVM:

$$M = \left\{ \frac{1}{2} | \uparrow_x \uparrow_x \rangle \langle \uparrow_x \uparrow_x |, \frac{1}{2} | \downarrow_x \downarrow_x \rangle \langle \downarrow_x \downarrow_x |, \frac{1}{2} | \uparrow_y \uparrow_y \rangle \langle \uparrow_y \uparrow_y |, \frac{1}{2} | \downarrow_y \downarrow_y \rangle \langle \downarrow_y \downarrow_y |, \frac{1}{2} | \uparrow_z \downarrow_z \rangle \langle \uparrow_z \downarrow_z | \right\}.$$  \hspace{1cm} (62)

This POVM cannot be realized by separate measurements on each particle, since the last term projects onto an entangled state.

For this POVM one calculates that $I_{ij}(M, \theta_i = 0) = \delta_{ij}$. Hence the left-hand side of Eq. (27) evaluates to $\sum_{ij} H^{-1}_{ij}(\theta_i = 0) I_{ij}(M, \theta_i = 0) = 3 > N(d - 1) = 2$. This shows that the optimal Fisher information is non additive.

5.6. Comparison with other Quantum Cramér-Rao bounds

An important question raised by bound (27) is how it compares to other quantum Cramér-Rao bounds obtained in the literature. In this respect, our most important result is that Eq. (27) is both a necessary and sufficient condition that $I(M, \theta)$ must satisfy when the dimensionality of the system $d$ equals 2 and the state is pure. This will be proven and discussed in detail in Sec. 7.

When $d > 2$, Eq. (27) is not a sufficient condition that $I(M, \theta)$ must satisfy. To see this let us compare Eq. (27) with the bound derived by Helstrom based on the SLD. This bound is the matrix inequality $I^N(M, \theta) \leq N H(\theta)$;
see Eq. (26). The comparison is most easily carried out by defining the matrix
\[ F = \frac{1}{N} H^{-\frac{1}{2}} I^N H^{-\frac{1}{2}} = \sum_{i=1}^{p} \gamma_i f_i \otimes f_i \]
where \( \gamma_i \) are the eigenvalues of \( F \) and \( f_i \) its eigenvectors. Helstrom’s bound can be reexpressed as \( \gamma_i \leq 1 \) for all \( i \), whereas bound (27) states that \( \sum_i \gamma_i \leq d - 1 \). From these expressions it results that the bound (27) is better than Helstrom’s bound for \( d = 2 \).

For \( d > 2 \) and \( p \leq d - 1 \), Helstrom’s bound is better than Eq. (27) as is seen by summing the inequalities \( \gamma_i \leq 1 \) too obtain \( \sum_i \gamma_i \leq p \). For \( p > d - 1 \), neither Helstrom’s bound nor bound (27) are better than the other.

Yuen and Lax [24] proposed another matrix bound based on an asymmetric logarithmic derivative. This bound is known to be worse than the bound based on the SLD in the case of one parameter, but it can be better, for some loss functions, in the case of two or more parameters. We have however not been able to make a detailed comparison between the bound based on the ALD and Eq. (27).

Although when \( d > 2 \), bound (27) is not a sufficient condition it can be complemented by additional constraints based on partial traces of \( H^{-1} I^N(M, \theta) \) which we now exhibit.

Consider a subset \( i = 1, \ldots, p' \) \( (p' < p) \) of the parameters. Let \( \rho_{i,i'} \) be the corresponding derivatives of \( \rho(\theta) \). Let us define the effective dimension \( d' \) of the space in which these parameters act at the point \( \theta^0 \) as follows. Let \( \Pi \) be a projector that commutes with \( \rho(\theta^0) \) \( ([\Pi, \rho(\theta^0)] = 0) \) and such that \( \rho_{i,i'}, i' = 1, \ldots, p' \) acts only within the eigenspace of \( \Pi \) (that is \( \Pi \rho_{i,i'} \Pi = \rho_{i,i'} \)). Then \( d' \) is the smallest dimension of the eigenspace of such a projector \( \Pi \) \( (d' = \text{Tr} \Pi) \). To be more explicit, let us reexpress the definition of \( d' \) in coordinates. First we diagonalize \( \rho(\theta^0) = \sum_k p_k |k\rangle \langle k| \). If some \( p_k \) are equal this can be done in many ways. The projector \( \Pi \) projects onto some of the eigenvectors of \( \rho: \Pi = \sum_{k=1}^{d'} |k\rangle \langle k| \). Next we write the operators \( \rho_{i,i'} \) in this basis: \( \rho_{i,i'} = \sum_{k,l=1}^{d'} (\rho_{i,i'})_{kl} |k\rangle \langle l| \), where the fact that the indices \( k, l \) go from one to \( d' \) expresses the fact that \( \rho_{i,i'} \) acts only within the eigenspace of \( \Pi \). Finally we choose the smallest such \( d' \).

We will show that
\[
\sum_{i',j'=1}^{p'} H_{i,j'}^{-1} I_{i,j'}^N(M, \theta^0) \leq N(d' - 1) .
\]

Before proving this result let us illustrate it by an example. Consider an unknown pure state in \( d \) dimensions. In the neighbourhood of a particular point we can parameterize the state by
\[
\psi = |1\rangle + (\theta_2 + i\eta_2) |2\rangle + \ldots + (\theta_d + i\eta_d) |d\rangle
\]
where the unknown parameters are $\theta_i$ and $\eta_i$, $i = 2, \ldots, d$. There are thus $2d - 2$ parameters. At the point $\theta = \eta = 0$, $H$ is diagonal in this parameterization: $H_{\theta_i\theta_j} = \delta_{ij}$, $H_{\eta_i\eta_j} = \delta_{ij}$, and $H_{\theta_i\eta_j} = 0$. Hence Eq. (27) takes the form

$$\sum_i I^N_{\theta_i\theta_i}(M, \theta = \eta = 0) + I^N_{\eta_i\eta_i}(M, \theta = \eta = 0) \leq N(d - 1).$$

(65)

But using (63) we also find the constraints

$$I^N_{\theta_i\theta_i}(M, \theta = \eta = 0) + I^N_{\eta_i\eta_i}(M, \theta = \eta = 0) \leq N, \quad i = 2, \ldots, d$$

(66)

which are stronger than Eq. (65) since they must hold separately, but by summing them one obtains Eq. (65).

The proof of Eq. (63) proceeds as in Sec. 5. First we can restrict ourselves to POVM’s whose elements are proportional to one dimensional projectors. Second we can restrict ourselves to the subspace $\Pi$ in evaluating Eq. (63). This follows from the inequality

$$I(M)^{i', j'} = \sum_\xi \frac{\text{Tr}(\rho_{i'}M_\xi) \text{Tr}(\rho_{j'}M_\xi)}{\text{Tr}(\rho M_\xi)}$$

$$= \sum_\xi \frac{\text{Tr}(\rho_{i'}\Pi M_\xi \Pi) \text{Tr}(\rho_{j'}\Pi M_\xi \Pi)}{\text{Tr}(\rho \Pi M_\xi \Pi) + \text{Tr}(\rho (1 - \Pi) M_\xi (1 - \Pi))}$$

$$\leq \sum_\xi \frac{\text{Tr}(\rho_{i'}\Pi M_\xi \Pi) \text{Tr}(\rho_{j'}\Pi M_\xi \Pi)}{\text{Tr}(\rho \Pi M_\xi \Pi)}.$$  

(67)

Note that equality in Eq. (67) holds when the measurement consists of one dimensional projectors and when the POVM decomposes into the sum of two POVM’s acting on the subspaces spanned by $\Pi$ and $1 - \Pi$ separately (i.e., the POVM elements $M_\xi = |\psi_\xi\rangle \langle \psi_\xi|$ must commute with $\Pi$ and $1 - \Pi$). Third, we can increase the number of parameters from $p'$ to $d'^2 - 1$. We then introduce exactly as in Eq. (46) a parameterization in which the $\rho_{i'}$ are particularly simple, but in place of Eq. (52) we use

$$\sum_{1 \leq m' \leq d'} c_{m'k'} c_{m'k'} = \delta_{k'k'} p_{k'} - \frac{p_{k'} p_{k'}}{\text{Tr}(\Pi \rho)}. \quad (68)$$

After these preliminary steps the left-hand side of Eq. (63) is calculated exactly as in Secs. 5.2, 5.3, and 5.4.
6. Dropping the Condition of Unbiased Estimators

6.1. Quantum van Trees inequality

In the previous section we proved a bound on the MQE of unbiased estimators $\hat{\theta}^N$ of $N$ copies of the quantum system $\rho(\theta)$ (with the additional condition that if $\rho$ is mixed the measurement should be separable). In this section we shall prove Theorems 4 and 5, that under additional conditions it is possible to drop the hypothesis that the estimator is unbiased.

The starting point for the results in this section is a Bayesian form of the Cramér-Rao inequality, the van Trees inequality [19], and in particular the multivariate form of the van Trees inequality proven in Ref. [7]. Adapted to the problem of estimating a quantum state, this inequality takes the following form. Let $\hat{\theta}^N$ be an arbitrary estimator of the parameter $\theta$ based on a measurement $M$ of the system $\rho^N(\theta)$. Suppose it has MQE matrix $V^N(\theta)$, and Fisher information matrix $I^N(M, \theta)$. Let $\lambda(\theta)$ be a smooth density supported on a compact region (with smooth boundary) of the parameter space, and suppose $\lambda$ vanishes on the boundary. By $E_\lambda$ we denote expectation over a random parameter value $\Theta$ with the probability density $\lambda(\theta)$. Let $C(\theta)$ and $D(\theta)$ be two $p \times p$ matrix valued functions of $\theta$, the former being symmetric and positive definite. Then the multivariate van Trees inequality reads

$$E_\lambda \text{tr} C(\Theta) V^N(\Theta) \geq \frac{(E_\lambda \text{tr} D(\Theta))^2}{E_\lambda \text{tr} C(\Theta)^{-1} D(\Theta) I^N(M, \Theta) D(\Theta)^\top + \tilde{I}(\lambda)},$$

(69)

where $^\top$ denotes the transpose of the matrix, and

$$\tilde{I}(\lambda) = \int d\theta \frac{1}{\lambda(\theta)} \sum_{ijkl} C_{ij}(\theta)^{-1} \partial_{\theta_k} \{D_{ik}(\theta) \lambda(\theta)\} \partial_{\theta_l} \{D_{jl}(\theta) \lambda(\theta)\}.$$  

(70)

As a first application of this inequality we shall prove Theorem 5, that is bound the minimum value averaged over $\theta$ of a quadratic cost function. Let $C(\theta)$ be the minimum value averaged over $\theta$ of a quadratic cost function. Consider the matrix $W_{opt}(\theta)$ that minimizes for each value of $\theta$ the cost $\text{tr} C(\theta) W(\theta)$ under the condition that $\text{tr} H(\theta)^{-1} W(\theta)^{-1} \leq d - 1$. One easily finds that

$$W_{opt} = \frac{\text{tr} \sqrt{H^{-1/2}C H^{-1/2}}}{d - 1} H^{-1/2} \sqrt{H^{1/2}C^{-1}H^{1/2}H^{-1/2}}$$

(71)

$$= \frac{\text{tr} \sqrt{C^{1/2}H^{-1}C^{1/2}}}{d - 1} C^{-1/2} \sqrt{C^{1/2}H^{-1}C^{1/2}C^{-1/2}}$$

(72)
and that
\[ \text{tr} CW_{opt} = \frac{\left( \text{tr} \sqrt{H^{-1/2}CH^{-1/2}} \right)^2}{d-1} = \frac{\left( \text{tr} \sqrt{C^{1/2}H^{-1}C^{1/2}} \right)^2}{d-1}. \] (73)

In Eq. (69) we choose \( D(\theta) = C(\theta)W_{opt}(\theta) \). Thus \( \text{tr} D(\theta) = \text{tr} C(\theta)W_{opt}(\theta) \) is given by Eq. (73). Note that
\[ D(\theta)^\top C(\theta)^{-1} D(\theta) = W_{opt}(\theta)C(\theta)W_{opt}(\theta) = \frac{\text{tr} C(\theta)W_{opt}(\theta)}{d-1} H(\theta)^{-1}. \] (74)

Thus
\[ \text{tr} D(\theta)^\top C(\theta)^{-1} D(\theta) I^N(M, \theta) = \frac{\text{tr} C(\theta)W_{opt}(\theta)}{d-1} \text{tr} H(\theta)^{-1} I^N(M, \theta) \leq N \text{tr} C(\theta)W_{opt}(\theta). \] (75)

Inserting these expressions into Eq. (69), one obtains
\[
E_\lambda \text{tr} C(\Theta)V^N(\Theta) \geq \frac{(E_\lambda \text{tr} C(\Theta)W_{opt}(\Theta))^2}{NE_\lambda \text{tr} C(\Theta)W_{opt}(\Theta) + \tilde{I}(\lambda)} \geq \frac{E_\lambda \text{tr} C(\Theta)W_{opt}(\Theta)^2}{N} - \frac{\alpha}{N^2}
\]
where
\[
\alpha = \frac{\tilde{I}(\lambda)}{E_\lambda \text{tr} C(\Theta)W_{opt}(\Theta)} \quad (76)
\]
is independent of \( N \). This proves that upon averaging over \( \theta \) it is impossible (for large \( N \)) to improve over the minimum cost [Eq. (30)].

6.2. Asymptotic version of the Cramér-Rao inequality

We now prove Theorem 4, that is an asymptotic version of our main inequality (28) which is valid at every point \( \theta \) and does not make the assumption of unbiased estimators. We must, however, slightly restrict the class of competing estimators since otherwise by the phenomenon of super-efficiency we can beat a given estimator at any specific value of the parameter, though we pay for this by bad behaviour closer and closer to the chosen value as \( N \) becomes larger.

The restriction on the class of estimators is that \( N \) times their mean quadratic error matrix must converge uniformly in a neighbourhood of the true value \( \theta^0 \) of \( \theta \) to a limit \( W(\theta) \), continuous at \( \theta^0 \). We assume that both \( W(\theta^0) \) and \( H(\theta^0) \) are nonsingular. Furthermore, we shall require some mild
smoothness conditions on \( H(\theta) \) in a neighbourhood of \( \theta^0 \): it must be continuous at \( \theta^0 \) with bounded partial derivatives with respect to the parameter in a neighbourhood of \( \theta^0 \). Note that imposing regularity conditions on \( H \) is natural since it corresponds to supposing that the \( \theta_i \) smoothly parameterize the allowed density matrices.

Suppose that, as \( N \to \infty \),

\[
NV^N(\theta) \to W(\theta)
\]

uniformly in \( \theta \) in a neighbourhood of \( \theta^0 \), with \( W \) continuous at \( \theta^0 \); write \( W^0 = W(\theta^0) \). Now in Eq. (69) let us make the following choices for the matrix functions \( C \) and \( D \):

\[
C(\theta) = W^0^{-1}H^{-1}(\theta)W^0^{-1}, \quad D(\theta) = W^0^{-1}H^{-1}(\theta).
\]

Then Eq. (69) (multiplied throughout by \( N \)) and (70) become

\[
E_{\lambda} \mathrm{tr} W^0^{-1}H^{-1}(\Theta)W^0^{-1} NV^N(\Theta) \geq \frac{(E_{\lambda} \mathrm{tr} W^0^{-1}H^{-1}(\Theta))^2}{\frac{1}{N}E_{\lambda} \mathrm{tr} H^{-1}I^N(M, \Theta) + \frac{1}{N} \tilde{I}(\lambda)} \geq \frac{(E_{\lambda} \mathrm{tr} W^0^{-1}H^{-1}(\Theta))^2}{(d - 1) + \frac{1}{N} \tilde{I}(\lambda)}
\]

(77)

and

\[
\tilde{I}(\lambda) = \int d\theta \frac{1}{\lambda(\theta)} \sum_{ijkl} H_{ij}(\theta)\partial_{\theta_k}\{H^{-1}_{ik}(\theta)\lambda(\theta)\} \partial_{\theta_l}\{H^{-1}_{jl}(\theta)\lambda(\theta)\},
\]

(78)

where we have used our central inequality (27) to pass to Eq. (77). Now suppose that the quantity (78) is finite (we will give conditions for that in a moment). By the assumed uniform convergence of \( NV^N \) to \( W \), upon letting \( N \to \infty \), Eq. (77) becomes

\[
E_{\lambda} \mathrm{tr} W^0^{-1}H^{-1}(\Theta)W^0^{-1} W(\Theta) \geq \frac{(E_{\lambda} \mathrm{tr} W^0^{-1}H^{-1}(\Theta))^2}{(d - 1)}. \quad \text{(79)}
\]

Now suppose the density \( \lambda \) in this equation (the probability density of \( \Theta \)) is replaced by an element \( \lambda^m \) in a sequence of densities, concentrating on smaller and smaller neighbourhoods of \( \theta^0 \) as \( m \to \infty \). Assume that \( H(\theta) \) is continuous at \( \theta^0 \). Recall our earlier assumption that \( W(\theta) \) is also continuous at \( \theta^0 \), with \( W^0 = W(\theta^0) \). Then taking the limit as \( m \to \infty \) of Eq. (79) yields

\[
\mathrm{tr} W^{-1}(\theta^0)H^{-1}(\theta^0) \geq (\mathrm{tr} W^{-1}(\theta^0)H^{-1}(\theta^0))^2/(d - 1)
\]
or the required limiting form of Eq. (27):
\[
\text{tr } W^{-1}(\theta^0)H^{-1}(\theta^0) \leq (d - 1).
\]

It remains to discuss whether it was reasonable to assume that \( \tilde{I}(\lambda^m) \) is finite (for each \( m \) separately). Note that this quantity only depends on the prior density \( \lambda \) and on \( H(\theta) \), where \( \lambda \) is one of a sequence of densities supported by smaller and smaller neighbourhoods of \( \theta^0 \). We already assumed that \( H(\theta) \) was continuous at \( \theta^0 \). It is certainly possible to specify prior densities \( \lambda^m \) concentrating on the ball of radius \( 1/m \), say, satisfying the smoothness assumptions in Ref. [7] and with, for each \( m \), finite Fisher information matrix
\[
\int d\theta \frac{1}{\lambda^m(\theta)} \partial_{\theta_k} \{\lambda^m(\theta)\} \partial_{\theta_l} \{\lambda^m(\theta)\}.
\]

Consideration of Eq. (78) then shows that it suffices further just to assume that \( \partial_{\theta_k} \{H^{-1}_{ik}(\theta)\} \) is, for each \( i, k \), bounded in a neighbourhood of \( \theta^0 \).

In conclusion we have shown that under mild smoothness conditions on \( H(\theta) \), the limiting mean quadratic error matrix \( W \) of a sufficiently regular but otherwise arbitrary sequence of estimators must satisfy the asymptotic version of our central inequality \( \text{tr } H^{-1}W^{-1} \leq d - 1 \).

7. Attaining the Cramér-Rao Bound in Two Dimensions

We shall now show that bounds (27), (29), and (31) are sharp in the case of pure states of spin-1/2 systems, and of separable measurements in the case of mixed states of spin-1/2 systems. In particular, in the limit of a large number of copies \( N \) any target scaled MQE matrix \( W \) that satisfies \( \text{tr } H^{-1}W^{-1} \leq 1 \) can be attained (provided \( W \) is non singular). We shall show this by explicitly constructing a measurement strategy that attains the bound. In Sec. 6 we have already shown that if \( \text{tr } H^{-1}W^{-1} > 1 \), then it cannot be attained.

7.1. Attaining the bound at a fixed point \( \theta^0 \)

The first step in the proof is to consider the case of one copy of the unknown state \( (N = 1) \) and fix a particular point \( \theta^0 \). Then we show that for any target information matrix \( G(\theta^0) \) that satisfies \( \text{tr } H^{-1}(\theta^0)G(\theta^0) \leq 1 \), we can build a measurement \( M = M^{\theta^0} \), in general depending on \( \theta^0 \), such that \( I(M^{\theta^0}, \theta^0) = G(\theta^0) \). In the next sections we shall show how to use this intermediate result to build a measurement and estimation strategy whose asymptotic MQE is equal to \( W(\theta) = G(\theta)^{-1} \) for all \( \theta \).
Let us first consider the case of pure states. At \( \theta^0 \), the state is \( |\psi^0\rangle \). We introduce a parameterization \( \theta_1, \theta_2 \) such that in the vicinity of \( |\psi^0\rangle \), the unknown state is

\[
|\psi(\theta)\rangle = |\psi^0\rangle + (\theta_1 + i\theta_2)|\psi^1\rangle.
\]

Thus the original point \( \theta^0 \) corresponds to the new \( \theta_1 = \theta_2 = 0 \). In this parameterization, \( H \) is proportional to the identity at \( \theta_1 = \theta_2 = 0 \): \( H_{\theta_1,\theta_2}(0) = H_{\theta_2,\theta_2}(0) = 1, H_{\theta_1,\theta_2}(0) = 0 \).

We now diagonalize the matrix \( G \). Thus there exist new parameters \( \theta_1' = \cos \lambda \theta_1 + \sin \lambda \theta_2 \) and \( \theta_2' = -\sin \lambda \theta_1 + \cos \lambda \theta_2 \) such that \( G_{\theta_1',\theta_2'}(0) = g_1 \geq 0, G_{\theta_1',\theta_2'}(0) = g_2 \geq 0, \) and \( G_{\theta_1',\theta_2'}(0) = 0 \).

In terms of the parameters \( \theta_1', \theta_2' \), the unknown state is written

\[
|\psi^0\rangle = |\psi^0\rangle + (\theta_1' + i\theta_2')|\psi^{1'}\rangle
\]

where \( |\psi^{1'}\rangle = e^{i\lambda}|\psi^1\rangle \).

The POVM \( M^{\theta^0} \) consists of measuring the observable \( |\psi^0\rangle\langle\psi^{1'}| + |\psi^{1'}\rangle\langle\psi^0| \) with probability \( g_1 \), of measuring the observable \( i(|\psi^0\rangle\langle\psi^{1'}| - \langle\psi^{1'}|\psi^0\rangle) \) with probability \( g_2 \), and of measuring nothing (or measuring the identity) with probability \( 1 - g_1 - g_2 \). It is straightforward to verify that the Fisher information at \( \theta^0 \) in a measurement of the POVM \( M^{\theta^0} \) is equal to \( G(\theta^0) \).

Let us now turn to the case of mixed states. We suppose that there are three unknown parameters. We use a parameterization in which \( \rho(\theta) = (1/2)(I + \theta \cdot \sigma) \), with \( \|\theta\| < 1 \). Without loss of generality we can suppose that \( \theta^0 = (0, 0, n) \), so that \( \rho(\theta^0) = (1/2 + n/2)|1\rangle\langle1| + (1/2 - n/2)|2\rangle\langle2| = \frac{1}{2}(I + n\sigma_z) \).

The tangent space at \( \rho \) is spanned by the Pauli matrices \( \rho_x = \sigma_x(= \frac{\rho_{12}^+}{2}) \), \( \rho_y = \sigma_y(= \frac{\rho_{12}^-}{2}) \), and \( \rho_z = \sigma_z(= \rho_{11}\sqrt{1 - n^2}) \), where in parentheses we have given the relation to the basis used in Sec. 5.3. In this coordinate system \( H(\theta^0) \) is diagonal with eigenvalues 1, 1, 1/\((1 - n^2) \).

Take any symmetric positive matrix \( G \) satisfying \( \text{tr} GH^{-1}(\theta^0) \leq 1 \). Define the matrix \( F = H^{-\frac{1}{2}}GH^{-\frac{1}{2}} = \sum_i \gamma_i f_i \otimes f_i \), where \( \gamma_i \) and \( f_i \) are the eigenvalues and eigenvectors of \( F \). The condition \( \text{tr} GH^{-1}(\theta^0) \leq 1 \) can then be rewritten \( \sum_i \gamma_i \leq 1 \). If we define \( g_i = H^{\frac{1}{2}} f_i \), then we can write \( G = \sum_i \gamma_i g_i \otimes g_i \). Denote \( m_i = g_i/\|g_i\| \).

Consider the measurement of the spin along the direction \( m_i \). This is the POVM consisting of the two projectors \( P_{+m_i} = \frac{1}{2}(I + m_i \cdot \sigma) \) and \( P_{-m_i} = \frac{1}{2}(I - m_i \cdot \sigma) \). The information matrix for this measurement is

\[
I(P_{\pm m_i})_{kl} = \sum_{\pm} \frac{\text{Tr}(P_{\pm m_i} \sigma_k) \text{Tr}(P_{\pm m_i} \sigma_l)}{\text{Tr}(P_{\pm m} \rho)} = \frac{(m_i)_k (m_i)_l}{(1 - n^2 (m_i)_z^2)}
\]

(82)
where \((m_i)_k\) is component \(k\) of vector \(m_i\). Therefore this information matrix is proportional to \(g_i \otimes g_i\). One verifies that it obeys \(\text{tr} H^{-1} I(P_{\pm m_i}) = 1\), as it must by our findings in Sec. 5 since the measurement is exhaustive, \(N = 1\), and \(p = d^2 - 1\). Therefore, the coefficient of proportionality is 1, and

\[
I(P_{\pm m_i}) = g_i \otimes g_i. \tag{83}
\]

We now combine such POVM’s to obtain the POVM whose elements are

\[
\gamma_1 P_{+m_1}, \gamma_1 P_{-m_1}, \gamma_2 P_{+m_2}, \gamma_2 P_{-m_2}, \gamma_3 P_{+m_3}, \gamma_3 P_{-m_3}, (1 - \gamma_1 - \gamma_2 - \gamma_3). \tag{84}
\]

The information matrix for this measurement is just the sum \(\gamma_1 I(P_{\pm m_1}) + \gamma_2 I(P_{\pm m_2}) + \gamma_3 I(P_{\pm m_3}) = \sum_i \gamma_i g_i \otimes g_i = G\). Thus POVM (84) attains the target information \(G\) at the point \(\theta^0\).

### 7.2. Attaining the bound for every \(\theta\) and arbitrary \(N\) by separable measurements

We now prove Theorem 7, that states that in the case of spin half particles we can attain the bound (29) for every \(\theta\). Give yourself a continuous matrix \(W(\theta)\), the target-scaled MQE matrix, satisfying Eq. (29) for every \(\theta\). Define \(G(\theta) = W(\theta)^{-1}\), the target-scaled information matrix, which therefore satisfies Eq. (27). We will show that there exists a separable measurement and an estimation strategy on \(N\) copies of the state \(\rho(\theta)\) such that the MQE matrix \(V_N^N\) of the estimator satisfies

\[
V_N^N(\theta)_{ij} = \text{E}_\theta((\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)) = \frac{W_{ij}(\theta)}{N} + o\left(\frac{1}{N}\right) \tag{85}
\]

for all \(\theta\). In fact this holds uniformly in \(\theta\) in a sufficiently small neighbourhood of any given point. This is proven by constructing explicitly a measurement and estimation strategy that satisfies Eq. (85), following the lines of Ref. [1].

The measurement and estimation strategy we propose is the following: first take a fraction \(N_0 = O(N^a)\) of the states, for some fixed \(0 < a < 1\), and on one-third of them measure \(\sigma_x\), on one third \(\sigma_y\) and on one-third \(\sigma_z\). One obtains from each measurement of \(\sigma_x\) the outcome \(\pm 1\) with probabilities \(\frac{1}{2}(1 \pm \theta_x)\), and similarly for \(\sigma_y, \sigma_z\). Using these data we make a first estimate of \(\theta\), call it \(\hat{\theta}\), for instance by equating the observed relative frequencies of \(\pm 1\) in the three kinds of measurement to their theoretical values. If the state is pure this determines a first estimate of the direction of polarization. If the
state is mixed it is possible that the initial estimate suggests that the Bloch vector lies outside the unit sphere. This only occurs with exponentially small probability (in $N_0$), and if this is the case the measurement is discarded. As discussed below this only affects the mean quadratic error by $o(1/N)$.

On the remaining $N' = N - N_0$ states we carry out the measurement $M = M^{\hat{\theta}}$ such that $I(M^{\hat{\theta}}, \hat{\theta}) = G(\hat{\theta})$, which we have just shown how to construct. Note that $I(M^{\hat{\theta}}, \theta) = G(\theta)$ is only guaranteed when $\theta$ is precisely equal to $\hat{\theta}$. Write $I(M, \theta; \hat{\theta})$ for the Fisher information about $\theta$, based on the measurement $M^{\hat{\theta}}$ optimal at $\hat{\theta}$, while the true value of the parameter is actually $\theta$. Given $\hat{\theta}$, each of the $N'$ second stage measurements represents one draw from the probability distribution $p(\xi | \theta; \hat{\theta}) = \text{Tr} M^{\hat{\theta}} \rho(\theta)$. We use the classical MLE based on this data only (with $\hat{\theta}$ fixed at its observed value) to estimate what is the value of $\theta$. Call this estimated value $\hat{\theta}$.

Let $\varepsilon > 0$ be fixed, arbitrarily small. Let $\theta^0$ denote the true value of $\theta$. For given $\delta > 0$ let $B(\theta^0, \delta)$ denote the ball of radius $\delta$ about $\theta^0$. Fix a convenient matrix norm $\| \cdot \|$. We have the exponential bound

$$\Pr\{\tilde{\theta} \in B(\theta^0, \delta)\} \geq 1 - Ce^{-DN_0\delta^2} \quad (86)$$

for some positive numbers $C$ and $D$ (depending on $\delta$). The reason we take $N_0$ proportional to $N^a$ for some $0 < a < 1$ is that this ensures that $1 - Ce^{-DN_0} = o(1/N)$.

Modern results [11] on the MLE $\hat{\theta}$ state that, under certain regularity conditions, the conditional MQE matrix of $\hat{\theta}$ satisfies (at $\theta = \theta^0$, and conditional on $\hat{\theta}$)

$$V^{N'}(\theta^0; \hat{\theta}) = \frac{I(M, \theta^0; \hat{\theta})^{-1}}{N'} + o\left(\frac{1}{N'}\right) \quad (87)$$

uniformly in $\theta^0$. However, for the next step in our argument this same result must be true uniformly in $\hat{\theta}$ for given $\theta^0$. This could be verified by careful reworking of the proof in Ref. [11]. Rather than doing this, in Secs. 7.3 and 7.4 we will explicitly calculate the conditional MQE matrix of our estimator, and show that it satisfies Eq. (87) uniformly in $\hat{\theta}$ in a small enough neighbourhood $B(\theta^0, \delta)$ of $\theta^0$. The “little $o$” in Eq. (87) refers to the chosen matrix norm.

We will also need that $I(M, \theta^0; \hat{\theta})^{-1}$ is continuous in $\hat{\theta}$ at $\hat{\theta} = \theta^0$, at which point it is equal to our construction the target-scaled MQE $W(\theta^0)$. This is also established in Sec. 7.3. Therefore, replacing if necessary $\delta$ by a smaller value, we can guarantee that $I(M, \theta^0; \hat{\theta})^{-1}$ is within $\varepsilon$ of $I(M, \theta^0; \theta^0)^{-1} = W(\theta^0)$ for all $\hat{\theta} \in B(\theta^0, \delta)$. 

If $\tilde{\theta}$ is outside the domain $B(\theta^0, \delta)$, then the norm of $V^{N'}(\theta^0; \tilde{\theta})$ is bounded by a constant $A$ since $\theta$ belongs to a compact domain.

Putting everything together we find that

$$\|N'V^N(\theta^0) - W(\theta^0)\| = \left\| \int \left( N'V^{N'}(\theta^0; \tilde{\theta}) - W(\theta^0) \right) dP(\tilde{\theta}) \right\|$$

$$\leq \int_{B(\theta^0, \delta)} \|N'V^{N'}(\theta^0; \tilde{\theta}) - W(\theta^0)\| dP(\tilde{\theta}) + AN'C'e^{-DN_0}$$

$$= \int_{B(\theta^0, \delta)} \|I(M, \theta^0; \tilde{\theta})^{-1} + o(1) - W(\theta^0)\| dP(\tilde{\theta}) + o(1) \leq \varepsilon + o(1) + o(1).$$

It follows since $N'/N \to 1$ as $N \to \infty$ that $\limsup_{N \to \infty} \|N'V(\theta^0) - W(\theta^0)\| \leq \varepsilon$. Since $\varepsilon$ was arbitrary, we obtain Eq. (85).

### 7.3. Analysis of the conditional mean quadratic error

We first consider the case of impure states, with the parameterization

$$\rho = \frac{1}{2}(I + \theta, \sigma), \quad \text{with} \quad \sum (\theta_i)^2 < 1. \quad (88)$$

where we have imposed that the state is never pure. This case turns out to allow the most explicit and straightforward analysis because the relation between the frequency of the outcomes and the parameters $\theta$ is linear. For other cases the analysis is more delicate, and is discussed in the next subsection. In general, smoothness assumptions will have to be made on the parameterization $\rho = \rho(\theta)$.

We suppose that $W(\theta)$ is non-singular and continuous in $\theta$. Consequently the $\gamma_i$ (defined in Sec. 7.1) depend continuously on $\theta$, and are all strictly positive at the true value $\theta^0$ of $\theta$.

Given the initial estimate, the second state measurement can be implemented as follows: for each of the $N' = N - N_0$ observations, independently of one another, with probability $\gamma_i$, measure the projectors $P_{\pm m_i}$, in other words, measure the spin observable $m_i, \sigma$. With probability $1 - \sum \gamma_i$, do nothing.

We emphasize that the $\gamma_i$ and $m_i$ all depend on the initial estimate $\tilde{\theta}$ through $W(\tilde{\theta})$ and $H(\tilde{\theta})$. In the following, all probability calculations are conditional on a given value of $\tilde{\theta}$.

For simplicity we will modify the procedure in the following two ways: first, rather than taking a random number of each of the three types of measurement, we will take the fixed (expected) numbers $[\gamma_i N']$ (and neglect
the difference between \([\gamma_i N']\) and \(\gamma_i N'\). Second, we will ignore the constraint \(\sum (\theta_i)^2 \leq 1\). These two modifications make the maximum likelihood estimator easier to analyze, but do not change its asymptotic MQE. Later we will sketch how to extend the calculations to the original constrained maximum likelihood estimator based on random numbers of measurements of each observable.

Now measuring \(m_i.\sigma\) produces the values \(\pm 1\) with probabilities \(p_{\pm i} = \frac{1}{2}(1 \pm \theta.m_i)\). Since our data consist of three binomially distributed counts and we have three parameters \(\theta_1, \theta_2\) and \(\theta_3\), the maximum likelihood estimator can be described, using the invariance of maximum likelihood estimators under 1–1 reparametrization, as follows: set the theoretical values \(p_{\pm i}\) equal to their empirical counterparts (relative frequencies of \(\pm 1\) in the \(\gamma_i N'\) observations of the \(i\)th spin) and solve the resulting three equations in three unknowns.

To be explicit, define \(\eta_i = 2p_{+i} - 1 = \theta.m_i\) and let \(\hat{\eta}_i\) be its empirical counterpart. Recall that \(m_i = g_i/\|g_i\|\) and \(g_i = H^{1/2}f_i\), where the \(f_i\) are the orthonormal eigenvectors of \(H^{-1/2}GH^{-1/2}\), and where \(H\) and \(G\) are \(H(\hat{\theta}), G(\hat{\theta})\), and \(\hat{\theta}\) is the preliminary estimate of \(\theta\). Then we can rewrite

\[
\eta_i = \theta \cdot m_i = \theta \cdot g_i/\|g_i\| = \theta \cdot H^{1/2}f_i/\|H^{1/2}f_i\| = (H^{1/2}\theta) \cdot f_i/\|H^{1/2}f_i\|,
\]

from which we obtain

\[
(H^{1/2}\theta) \cdot f_i = \|H^{1/2}f_i\|\eta_i
\]

and hence

\[
\theta = H^{-1/2} \sum_i \|H^{1/2}f_i\|\eta_i f_i.
\]

The same relation holds between \(\hat{\theta}\) and \(\hat{\eta}_i\) and yields the sought for expression for \(\hat{\theta}\) in terms of the empirical relative frequencies.

Observing that \(\hat{\eta}_i\) are independent with variance \(4p_{+m_i}p_{-m_i}/(\gamma_i N') = (1 - (\theta m_i)^2)/(\gamma_i N')\), the MQE matrix of \(\hat{\theta}\), conditional on the preliminary estimate \(\hat{\theta}\), is

\[
V^{N'}(\theta^0; \hat{\theta}) = \frac{1}{N'} \sum i \frac{1}{\gamma_i} \left(1 - \frac{(\theta^0 \cdot H^{1/2}f_i)^2}{\|H^{1/2}f_i\|^2}\right) \|H^{1/2}f_i\|^2 H^{-1/2}(f_i \otimes f_i)H^{-1/2}.
\]
There is no $o(1/N')$ term here, so we do not have to check uniform convergence: the limiting value is attained exactly. Actually we cheated by replacing $\lfloor \gamma_i N' \rfloor$ by $\gamma_i N'$. This does introduce a $o(1/N')$ error into Eq. (89) uniformly in a neighbourhood of $\theta^0$ in which $\gamma_i$, which depend on $\hat{\theta}$, are bounded away from zero, and $H$ and its inverse are bounded.

One may verify that Eq. (89) reduces to $W(\theta^0)/N'$ at $\hat{\theta} = \theta^0$ (indeed at $\theta^0 = \hat{\theta}$, $(\hat{\theta} \cdot H_{1/2} f_i)^2 = \frac{n^2 f_i^2}{1-n}$ and $\|H_{1/2} f_i\|^2 = \frac{1-n^2+n^2 f_i^2}{1-n^2}$). But this computation is really superfluous since, at this point, we are computing the MQE of the maximum likelihood estimator based on a measurement with, by our construction, Fisher information equal to the inverse of $W(\theta^0)$. (The modifications to our procedure do not alter the Fisher information). The two quantities must be equal by the classical large sample results for the maximum likelihood estimator.

We finally need to show the continuity in $\hat{\theta}$ at $\hat{\theta} = \theta^0$ of $N'$ times the quantity in Eq. (89). This is evident if $\gamma_i$ are all different at $\theta^0$. Both the eigenvalues and the eigenvectors of $H^{-1/2} G H^{-1/2}$ are then continuous functions of $\hat{\theta}$ at $\theta^0$. However, there is a potential difficulty if some $\gamma_i$ are equal to one another at $\hat{\theta} = \theta^0$. In this case, the eigenvectors $f_i$ are not continuous functions of $\hat{\theta}$ at this point, and not even uniquely defined there. We argue as follows that this does not destroy continuity of the mean quadratic error. Consider a sequence of points $\hat{\theta}^n$ approaching $\theta^0$. This generates a sequence of eigenvectors $f_i^n$ and eigenvalues $\gamma_i^n$. The eigenvalues converge to the $\gamma_i$ but the eigenvectors need not converge at all. However by compactness of the set of unit vectors in $\mathbb{R}^3$, there is a subsequence along which the eigenvectors $f_i^n$ converge; and they must converge to a possible choice of eigenvectors at $\theta^0$. Thus along this subsequence the mean quadratic error (89) does converge to a limit given by the same formula evaluated at the limiting $f_i$ etc. But this limit is equal by construction to the inverse of the target information matrix $G(\theta)$. A standard argument now shows that the limiting mean quadratic error is continuous at $\hat{\theta} = \theta^0$. The MQE of $\hat{\theta}$ given $\hat{\theta}$ (times $N'$) therefore converges uniformly in a sufficiently small neighbourhood of $\theta^0$ to a limit continuous at that point, and is equal to $W(\theta^0)$ there.

In our derivation of Eq. (85) we required the parameter and its estimator to be bounded. By dropping the constraint on the length of $\theta$ we have inadvertently lost this property. Suppose we replace our modified estimator $\hat{\theta}$ by the actual maximum likelihood estimator respecting the constraint. The two only differ when the unconstrained estimator lies outside the unit sphere; but this event only occurs with an exponentially small probability,
uniformly in $\tilde{\theta}$, provided $\gamma_i$ are uniformly bounded away from 0 in the given neighbourhood of $\theta^0$. From this it can be shown that the mean quadratic error is altered by an amount $o(1/N')$ uniformly in $\tilde{\theta}$.

If we had worked with random numbers of measurements of each spin variable, when computing the mean quadratic error we would first have copied the computation above conditional on the numbers of measurements, say $X_i$, of each spin $m_i$. These numbers are binomially distributed with parameter $N'$ and $\gamma_i$. The conditional mean quadratic error would be the same as the expression above but with $1/(\gamma_i N')$ replaced by $1/X_i$ (and special provision taken for the possible outcome $X_i = 0$). So to complete the argument we must show that $E(1/X_i) = 1/(\gamma_i N') + o(1/N')$ uniformly in $\tilde{\theta}$. This can also be shown to be true, using the fact that $X_i/N'$ only differs from its mean by more than a fixed amount with exponentially small probability as $N' \to \infty$, and we restrict attention to $\tilde{\theta}$ in a neighbourhood of $\theta^0$, where $\gamma_i$ are bounded away from zero. Inspection of our argument shows that the convergence of the mean quadratic error is uniform in $\theta^0$, as long as we keep away from the boundary of the parameter space.

By the convergence of the normalized binomial distribution to the normal distribution, the representation of the estimator we gave above also shows that it is asymptotically normally distributed with asymptotic covariance matrix equal to the target covariance matrix $W$. Moreover, if $X$ has the binomial $(n, p)$ distribution, then $n^{1/2}(X/n - p)$ converges in distribution to the normal with mean zero and variance $p(1 - p)$, uniformly in $p$. Thus the convergence in distribution is also uniform in $\theta^0$ as long as we keep away from the boundary of the parameter space.

7.4. Conditional mean quadratic error for other models

The preceding subsection gave a complete analysis of the mean quadratic error, given the preliminary estimate $\tilde{\theta}$ for the 3 unknown parameters $\theta_j$, of parameterization (88). We shall first analyze the mean quadratic error when the unknown parameters are functions $\phi_i(\theta_j)$ of the parameters $\theta_j$. We shall then consider the important case when the state is pure and depends on two unknown parameters, and finally the case when the state is pure or mixed and depends on one unknown parameter, or is mixed and depends on two unknown parameters.

Our first result is that if the change of parameters $\phi_i(\theta_j)$ is locally $C^1$, then the MQE matrix of the $\phi_i$ is obtained from the MQE of the $\theta_j$ by the Jacobian $\partial \phi_i/\partial \theta_j$ except eventually at isolated points. This follows from the
fact that under a smooth (locally $C^1$) parameterization, the $\delta$ method (first-order Taylor expansion) allows us to conclude a uniform convergence of the probability distribution of $\sqrt{N}(\hat{\phi}^N - \phi)$ to a normal limit with the target mean quadratic error. If the $\phi_i$ and their derivatives $\partial\phi_i/\partial\theta_j$ are bounded, then this proves our claim. If there are points where $\phi_i$ or their derivatives $\partial\phi_i/\partial\theta_j$ are infinite, then convergence in distributions not necessarily imply convergence of moments. However, a truncation device allows one to modify the estimate $\hat{\phi}$, replacing it by 0 if any component is larger than $cN^a$ for given $c$ and $a$ [use the method of Lemma II.8.2 [11], together with the exponential inequality (86) for the multinomial distribution]. With this minor modification one can show (uniform in $\phi$ in a neighbourhood of $\phi^0$) convergence of the moments of the corresponding $\sqrt{N}(\hat{\phi}-\phi)$ to the moments of its limiting distribution, and hence achievement of the bound in the sense of Theorem 4. In particular, if the parameter $\phi$ is bounded then the truncation is superfluous.

Now turn to the pure state analog of model (88). Obtain a preliminary estimate of the location of $\rho$ on the surface of the Poincaré sphere using the same method as in the mixed case, but always projecting onto the surface of the sphere. Next, after rotation to transform the preliminary estimate into “spin-up,” reparameterize to $\rho = \frac{1}{2}(1 + \phi \cdot \sigma)$, where the parameters to be estimated are $(\phi_1, \phi_2) = (\theta'_1, \theta'_2)$ of the parameterization (81) while $\phi_3 = \sqrt{(1 - \phi_1^2 - \phi_2^2)}$. The preliminary estimate is at $\phi_1 = \phi_2 = 0$. The optimal measurement at this point according to Sec. 7.1 consists of measurements of the spins $\sigma_1$ and $\sigma_2$ on specified proportions of the remaining copies. The resulting estimator of the parameter $(\phi_1, \phi_2)$ is a linear function of binomial counts, and hence its mean quadratic error can be studied exactly as in Sec. 7.3. Then we must transfer back to the originally specified parameterization, for instance polar coordinates. This is done as in the preceding paragraph. If the transformation is locally $C^1$, then uniform convergence in distribution to the normal law also transfers back; there is also a convergence of the mean quadratic error if the original parameter space is bounded. Otherwise a truncation might be necessary. In any case, we can exhibit a procedure optimal in the sense of Theorem 4.

It remains to consider one- and two-dimensional sub-models of the full mixed model, and one-dimensional sub-models of the full pure model. We suppose that the model specifies a smooth curve or surface in the interior of the Poincaré sphere, or a smooth curve on its surface; smoothly parameterized by a one- or two-dimensional parameter as appropriate. The first stage of the procedure is just as before, finishing in projection of an estimated
density matrix into the model. Then we reparameterize locally, augmenting the dimension of the parameter to convert the model into a full mixed or pure model, respectively. The target information for the extra parameters is zero. Compute as before the optimal measurement at this point. Because of the zero values in the target information matrix, there will be zero eigenvalues $\gamma_i$ in the computation of Sec. 7.1. Thus the optimal measurement will involve specified fractions of measurement of spin in the same number of directions as the dimension of the model. Compute the maximum likelihood estimator of the original parameters based on this data. If the parameterization is smooth enough, the estimator will yet again achieve the bound of Theorem 4.

8. Conclusions and Open Questions

In this paper we have solved some of the theoretical problems that arise when trying to estimate the state of a quantum system of which one possesses a large number of copies. This constitutes a preliminary step toward solving the question with which Helstrom concluded his book [8]: “... mathematical statisticians are often concerned with asymptotic properties of decision strategies and estimators, ... When the parameters of a quantum density operator are estimated on the basis of many observations, how does the accuracy of the estimates depend on the number of observations as that number grows very large? Under what conditions have the estimates asymptotic normal distributions? Problems such as these, and still others that doubtless will occur to physicists and mathematicians, remain to be solved within the framework of the quantum-mechanical theory.”

In the case of pure states of spin-1/2 particles, the problem has been solved. The key result is that in the limit of large $N$, the variance of the estimate is bounded by Eq. (28), and the bound can be attained by separate von Neumann measurements on each particle.

In the case of mixed states of spin-1/2 particles the state estimation problem for large $N$ has been solved if one restricts oneself to separable measurements. However, if one considers nonseparable measurements, then one can improve the quality of the estimate, which shows that the Fisher information, which in classical statistics is additive, is no longer so for quantum state estimation.

For the case of mixed states of spin-1/2 particles, or for higher spins, we do not know what the “outer” boundary of the set of (rescaled) achievable Fisher information matrices based on arbitrary (nonseparable) measure-
ments of $N$ systems looks like. We have some indications about the shape of this set (see Sec. 5.6), and we know that it is convex and compact.

Acknowledgments

S.M. thanks Utrecht University, The Netherlands, where part of this work was carried out. He is research associate of the Belgium National Research Fund. R.D.G. thanks the generous hospitality of the Department of Mathematics and Statistics, University of Western Australia.

References


Additional Note by Author
The paper (*), Gill (2001), presents the same results as Gill & Massar (2002), but is written for mathematical statisticians rather than for physicists. It contains a deeper mathematical discussion of alternative, precise, definitions of asymptotic optimality, and of the regularity conditions needed to guarantee results of various type (e.g., optimality of mean square error of limit, versus optimal limit of mean square error; problem of phenomenon of super-efficient estimators). An important innovation of these papers was the use of the van Trees inequality, a Bayesian version of Cramér-Rao, in order to deal rigorously with these issues.


§Additional author’s note: This paper exists only on quant-ph, not in journal form. It reviews various analytical (i.e., assuming no statistical error) approaches to state reconstruction for finite dimensional systems, including some approaches of the the author, which can be found (via quant-ph) in some other, published, papers.
PART III

Quantum Cramér-Rao Bound in Pure States Model

Chap. 17: A. Fujiwara and H. Nagaoka
“Quantum Fisher metric and estimation for pure state models” . . . . 220

Chap. 18: A. Fujiwara
“Geometry of quantum estimation theory” . . . . . . . . . . . . . . . . . . 229

Chap. 19: A. Fujiwara and H. Nagaoka
“An estimation theoretical characterization of coherent states” . . . . 287

Chap. 20: K. Matsumoto
“A geometrical approach to quantum estimation theory” . . . . . . . 305
CHAPTER 16

Introduction to Part III

1. Quantum Cramér-Rao Bound in Pure States Model

As mentioned in chapter 7, the asymptotic theory is one of the central issues in quantum statistical inference. However, the pure state case cannot be treated just as a special case of the mixed case. As is well known, a pure density matrix has no inverse. This property causes several difficulties in quantum estimation as well as other related topics. However, it has been found that this property of pure states simplifies the calculation of the CR bound. In the following, we review the history of these results.

In Chap. 7, we mentioned that Helstrom [II-1] solved an optimization problem in the one-parameter case. The SLD is not uniquely defined in the pure states case. On the other hand, Holevo uniquely defined the SLD as a vector in the the quantum $L^2$ space in a general setting (Chapters II and VI of Holevo [0-2]). Fujiwara and Nagaoka [Chap. 17] developed essentially the same argument for the pure state case in particular, and expressed the quantum $L^2$ space as a quotient space, concretely. They also calculated SLD Fisher information in the pure state case (Also see Fujiwara [III-1]).

Moreover, another difficulty is caused by the non-invertibility. As is mentioned in Chap. 7, Holevo derived the CR bound of the quantum Gaussian state family based on RLD Fisher information matrix. However, the RLD diverges, i.e., can not be defined in the pure state case. Thus, we cannot directly apply Holevo’s method to the family of pure coherent states while this family is the noiseless limit of the quantum Gaussian state family. In order to avoid this difficulty, Fujiwara and Nagaoka [III-2] invented another method and calculated the CR bound in this case. Their method can be applied to a certain two-parameter pure states family called the two-parameter coherent model, which contains the two-parameter pure states family in the quantum two-level system. Applying their method to the quantum two-level system, they provided a simpler proof than Nagaoka’s [Chap. 8] proof in this case. Combining both papers [Chap. 17, III-2], they pointed out the difference between the structures of SLD and RLD in the pure state case. Indeed, such a difficulty appears in the non-invertible case. Luati [III-3]
discussed POVMs achieving the optimal Fisher information when the rank of the density matrix is two.

Furthermore, Fujiwara extended their result to the multi-parameter coherent model and calculated the CR bound of this model in his doctoral thesis [Chap. 18]. Since his thesis contains reviews of previous results, e.g., the paper [III-2] etc., as well as several original results, it is quite introductory. Following this result, Fujiwara and Nagaoka [Chap. 19] simplified its proof. Moreover, Matsumoto invented a more general approach that could be applied to any pure states model [Chap. 20, III-5]. Using this method, he solved this problem in any two-parameter pure states family. His result covers all known results regarding the CR bound in the pure state case.

Next, we focus on the possibility of reducing the CR bound by the use of a quantum collective measurement on the tensor product space. It is easily checked by Matsumoto’s characterization of CR bounds in the pure states model, that the CR bound of the $n$-fold tensor product system in pure states model equals $\frac{1}{n}$ times that for one single system. This is because the CR bound can be defined only by the inner products between the vectors $\frac{\partial \psi(\theta)}{\partial \theta} - \langle \psi(\theta), \frac{\partial \psi(\theta)}{\partial \theta} \rangle \psi(\theta)$, where $\psi(\theta)$ is the state vector. This fact does not appear in Matsumoto [Chap. 20], but does in Matsumoto [III-5].

As mentioned in Chap. 7, Hayashi and Matsumoto [Chap. 12] and Gill and Massar [Chap. 15] showed that the first-order coefficient of the optimal error equals the limit of $n$ times the CR bound of the $n$-fold tensor product system. Combining it with the above statement, we can verify that the optimal error can be, asymptotically to the first order, attained by an adaptive measurement.

2. Geometrical Study of Quantum Estimation

It is necessary to discuss other problems apart from calculating the CR bound.

Since information geometry succeeded in geometrically characterizing several statistical properties in classical statistics [II-12], it is a natural problem to geometrically characterize quantum states families from an estimation-theoretical viewpoint. The first success of this direction is a quantum extension of the concept of ‘exponential family’ by Nagaoka [Chap. 9]. Nagaoka [Chap. 10, III-4] also considered quantum extension of exponential connection. Based on results, Fujiwara and Nagaoka [Chap. 17] and Fujiwara [Chap. 18] obtained remarkable results of geometrical aspects of quantum estimation theory as well as CR bounds of coherent model. They
characterized the quantum exponential family in the pure states case in a different form, and obtained an insightful relation with the time-energy uncertainty relation. Moreover, Matsumoto focused on Uhlmann geometry and Berry phase, and discussed their relation with quantum estimation theory [Chap. 20]. He also discussed the uncertainty principle from the viewpoint of quantum estimation theory [III-6].

Further Reading


CHAPTER 17

Quantum Fisher Metric and Estimation for Pure State Models

Akio Fujiwara and Hiroshi Nagaoka

Abstract. A statistical parameter estimation theory for quantum pure state models is presented. We first investigate the basic framework of pure state estimation theory and derive quantum counterparts of the Fisher metric. We then formulate a one-parameter estimation theory, based on the symmetric logarithmic derivatives, and clarify the differences between pure state models and strictly positive models.

1. Introduction

A quantum statistical model is a family of density operators $\rho_\theta$ defined on a certain separable Hilbert space $\mathcal{H}$ with finite-dimensional real parameters $\theta = (\theta^i)_{i=1}^n$ which are to be estimated statistically. In order to avoid singularities, the conventional quantum estimation theory [7, 8] has been often restricted to models that are composed of strictly positive density operators. It was Helstrom [6] who successfully introduced the symmetric logarithmic derivative for the one-parameter estimation theory as a quantum counterpart of the logarithmic derivative in the classical estimation theory. The right logarithmic derivative is another successful counterpart introduced by Yuen and Lax [15] in the expectation parameter estimation theory for quantum Gaussian models, which provided a theoretical background of optical communication theory. Quantum information theorists have also kept away from degenerate states, such as pure states, for mathematical convenience [12]. Indeed, the von Neumann entropy cannot distinguish the pure states, and the relative entropy diverge. In this paper, however, we try to construct an estimation theory for pure state models, and clarify the differences between the pure state case and the strictly positive state case.

2. Preliminaries

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \psi | \varphi \rangle$ for every $\psi, \varphi \in \mathcal{H}$. Further, let $\mathcal{L}$ and $\mathcal{L}_{sa}$ are, respectively, the set of all (bounded) linear operators and all self-adjoint operators on $\mathcal{H}$. Given a possibly degenerate density operator $\rho$, we define sesquilinear forms on $\mathcal{L}$:

\begin{align*}
(A, B)_\rho &= \text{Tr} \rho BA^*, \\
\langle A, B \rangle_\rho &= \frac{1}{2} \text{Tr} \rho (BA^* + A^*B),
\end{align*}

where $A, B \in \mathcal{L}$. These are pre-inner products on $\mathcal{L}$, i.e., possess all properties of inner products except that $(K, K)_\rho$ and $\langle K, K \rangle_\rho$ may be equal to zero for a nonzero $K \in \mathcal{L}$. Note that the Schwarz inequality also holds for pre-inner product. The forms $(\cdot, \cdot)_\rho$ and $\langle \cdot, \cdot \rangle_\rho$ become inner products if and only if $\rho > 0$. If rank $\rho = 1$ or equivalently $\rho^2 = \rho$, $\rho$ is called pure. The following lemmas are fundamental.

**Lemma 1:** Suppose $\rho$ is pure. Then the following three conditions for linear operators $K \in \mathcal{L}$ are equivalent.

\begin{enumerate}
  \item $(K, K)_\rho = 0,$
  \item $\rho K = 0,$
  \item $\text{Tr} \rho K = 0$ and $\rho K + K^* \rho = 0$.
\end{enumerate}

**Proof:** Let us express as $\rho = |\psi\rangle \langle \psi|$ where $|\psi\rangle$ is a normalized vector in $\mathcal{H}$. Then the following equivalent sequence

\begin{align*}
(K, K)_\rho = 0 &\iff \langle \psi | K K^* | \psi \rangle = 0 \iff \langle \psi | K = 0 \iff |\psi\rangle \langle \psi | K = 0
\end{align*}

yield (i)$\iff$(ii). Further, (ii)$\Rightarrow$(iii) is trivial. (iii)$\Rightarrow$(ii) is shown as follows. Operating $\langle \psi |$ from the left to the assumption

\begin{align*}
|\psi\rangle \langle \psi | K + K^* |\psi\rangle \langle \psi | = 0,
\end{align*}

and invoking another assumption $\text{Tr} \rho K = 0 \iff \langle \psi | K^* | \psi \rangle = 0$, we have

\begin{align*}
0 = \langle \psi | \psi \rangle \langle \psi | K + \langle \psi | K^* | \psi \rangle \langle \psi | \psi \rangle = \langle \psi | K.
\end{align*}

Therefore $|\psi\rangle \langle \psi | K = 0$. 

**Lemma 2:** Suppose $\rho$ is pure. Then the following three conditions for linear operators $K \in \mathcal{L}$ are equivalent.

\begin{enumerate}
  \item $\langle K, K \rangle_\rho = 0,$
  \item $\rho K = \rho K^* = 0,$
\end{enumerate}
(iii) $\text{Tr} \rho K = 0$, $\rho K + K^* \rho = 0$, and $\rho K^* + K \rho = 0$.

Proof: This is a straightforward consequence of Lemma 1 since $\langle K, K \rangle_\rho = \frac{1}{2}[(K, K)_\rho + (K^*, K^*)_\rho]$.

Lemma 3: Suppose $\rho$ is pure. Then the following three conditions for self-adjoint operators $K \in \mathcal{L}_{sa}$ are equivalent.

(i) $\langle K, K \rangle_\rho = 0$,
(ii) $\rho K = 0$,
(iii) $\rho K + K \rho = 0$.

Proof: Straightforward by setting $K = K^*$ in Lemma 2.

Denote by $\mathcal{K}(\rho)$ the set of linear operators $K \in \mathcal{L}$ satisfying $\langle K, K \rangle_\rho = 0$, which are called the kernel of the pre-inner product $\langle \cdot, \cdot \rangle_\rho$. Also denote by $\mathcal{K}_{sa}(\rho)$ the set of self-adjoint operators $K \in \mathcal{L}_{sa}$ satisfying $\langle K, K \rangle_\rho = 0$, which are called the kernel of the pre-inner product $\langle \cdot, \cdot \rangle_\rho$. Note $\mathcal{K}_{sa}(\rho) = K(\rho) \cap \mathcal{L}_{sa}$.

3. Quantum Fisher Metric

Suppose we are given an $n$-parameter pure state model:

$$S = \{ \rho_\theta : \rho_\theta^* = \rho_\theta, \text{Tr} \rho_\theta = 1, \rho_\theta^2 = \rho_\theta, \theta \in \Theta \subset \mathbb{R}^n \}.$$ (3)

We define a family of quantum analogues of the logarithmic derivative by

$$\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2}[\rho_\theta L_{\theta,j} + L_{\theta,j}^* \rho_\theta], \quad \text{Tr} \rho_\theta L_{\theta,j} = 0.$$ (4)

For instance,

$$\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2}[\rho_\theta L_{\theta,j}^S + L_{\theta,j}^S \rho_\theta], \quad L_{\theta,j}^S = L_{\theta,j}^*$$ (5)

defines the symmetric logarithmic derivative (SLD) $L_{\theta,j}^S$ introduced by Helstrom [6]. Furthermore, since every pure state model is written in the form $\rho_\theta = U_\theta \rho_0 U_\theta^*$, where $U_\theta$ is unitary, we have another useful logarithmic derivative

$$\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2}[\rho_\theta L_{\theta,j}^A - L_{\theta,j}^A \rho_\theta], \quad \text{Tr} \rho_\theta L_{\theta,j}^A = 0, \quad L_{\theta,j}^A^* = -L_{\theta,j}^A,$$ (6)

which may be called the anti-symmetric logarithmic derivative (ALD). Indeed, the ALD is closely related to the local generator $A_{\theta,j} = -i(\partial U_\theta/\partial \theta^j)U_\theta^*$ of the unitary $U_\theta$ such as $L_{\theta,j}^A = -2i(A_{\theta,j} - \text{Tr} \rho_\theta A_{\theta,j})$. 

Thus, (4) defines a certain family of logarithmic derivatives [10]. Denote by $T(\rho_\theta)$ the linear span (over $\mathbb{R}$) of logarithmic derivatives which satisfy (4).

**Lemma 4:** Suppose $\rho_\theta$ is pure and an arbitrary linear operator $A \in \mathcal{L}$ is given. Then all the quantities $(A, L_{\theta,j})_{\rho_\theta}$ are identical for every logarithmic derivatives $L_{\theta,j} \in T(\rho_\theta)$ which correspond to the same direction $\theta^j$.

**Proof:** Take any logarithmic derivatives $L_{\theta,j}$ and $L'_{\theta,j}$ which correspond to the same $\theta^j$, and denote $K = L_{\theta,j} - L'_{\theta,j}$. Then, from (4), $K$ satisfies condition (iii) in Lemma 1. Therefore $(K,K)_{\rho_\theta} = 0$ holds. This and the Schwarz inequality

$$|(A,K)_{\rho_\theta}|^2 \leq (A,A)_{\rho_\theta}(K,K)_{\rho_\theta},$$

lead to $(A,K)_{\rho_\theta} = 0$ for all $A \in \mathcal{L}$.

From Lemma 4, we can define uniquely the complex Fisher information matrix $J_\theta$ for the family of logarithmic derivatives (4) whose $(j,k)$ entry is $(L_{\theta,j}, L_{\theta,k})_{\rho_\theta}$.

The SLD is also not uniquely determined for pure state models. Denote by $T^S(\rho_\theta)$ the linear span (over $\mathbb{R}$) of SLDs which satisfy (5).

**Lemma 5:** Suppose $\rho_\theta$ is pure and an arbitrary self-adjoint operator $A \in \mathcal{L}_{sa}$ is given. Then all the quantities $(A, L^S_{\theta,j})_{\rho_\theta}$ are identical for every SLD $L^S_{\theta,j} \in T^S(\rho_\theta)$ which corresponds to the same direction $\theta^j$.

**Proof:** By using Lemma 3, it is proved in the same way as Lemma 4.

From Lemma 5, we can define uniquely the real Fisher information matrix $J^S_\theta$ for the family of SLDs (5) whose $(j,k)$ entry is $\langle L^S_{\theta,j}, L^S_{\theta,k}\rangle_{\rho_\theta}$, which is called the SLD–Fisher information matrix. The above results are summarized by the following theorem.

**Theorem 6:** Suppose $\{\rho_\theta\}$ is a pure state model. Then the complex Fisher information matrix $J_\theta = [(L_{\theta,j}, L_{\theta,k})_{\rho_\theta}]$ and the SLD–Fisher information matrix $J^S_\theta = [\langle L^S_{\theta,j}, L^S_{\theta,k}\rangle_{\rho_\theta}]$ are uniquely determined on the quotient spaces $T(\rho_\theta)/K(\rho_\theta)$ and $T^S(\rho_\theta)/K_{sa}(\rho_\theta)$, respectively. They are related by $J^S_\theta = \text{Re} J_\theta$. The $(j,k)$ entries of $J_\theta$ and $J^S_\theta$ become

$$(J_\theta)_{jk} = 4 \text{Tr} \rho_\theta (\partial_k \rho_\theta)(\partial_j \rho_\theta),$$

and

$$(J^S_\theta)_{jk} = 2 \text{Tr} (\partial_j \rho_\theta)(\partial_k \rho_\theta),$$
respectively, where \( \partial_j = \partial/\partial \theta^j \). In particular, \( J^S_\theta \) is identical, up to a constant factor, to the Fubini–Study metric.

**Proof:** We only need to prove (8). Differentiating \( \rho_\theta = \rho^2_\theta \),

\[
\partial_j \rho_\theta = (\partial_j \rho_\theta) \rho_\theta + \rho_\theta (\partial_j \rho_\theta).
\]

(9)

This relation indicates that \( 2\partial_j \rho_\theta \) is a representative of the SLD. Then

\[
(J^S_\theta)_{jk} = \langle 2\partial_j \rho_\theta, 2\partial_k \rho_\theta \rangle_{\rho_\theta} = 2 \text{Tr} \rho_\theta [(\partial_j \rho_\theta)(\partial_k \rho_\theta) + (\partial_k \rho_\theta)(\partial_j \rho_\theta)].
\]

(10)

Further, multiplying \( \rho_\theta \) with (9), we have

\[
\rho_\theta (\partial_j \rho_\theta) \rho_\theta = 0.
\]

(11)

Therefore, by using (9) and (11),

\[
(\partial_j \rho_\theta)(\partial_k \rho_\theta) = \left([\partial_j \rho_\theta] \rho_\theta + \rho_\theta (\partial_j \rho_\theta)\right) \left([\partial_k \rho_\theta] \rho_\theta + \rho_\theta (\partial_k \rho_\theta)\right)
\]

\[
= (\partial_j \rho_\theta) \rho_\theta (\partial_k \rho_\theta) + \rho_\theta (\partial_j \rho_\theta)(\partial_k \rho_\theta) \rho_\theta.
\]

This, along with (10), leads to relation (8). Denoting \( \rho_\theta = \langle \psi | \psi \rangle \)

\[
\text{Tr}(\partial_j \rho_\theta)(\partial_k \rho_\theta) = 2 \text{Re} \langle \partial_j \psi | \partial_k \psi \rangle + \langle \psi | \partial_j \psi \rangle \langle \psi | \partial_k \psi \rangle,
\]

which is identical to the Fubini–Study metric \([1, 9]\).

The Fubini–Study metric is known as a gauge invariant metric on a projective Hilbert space \([13]\). Theorem 6 gives another meaning of the Fubini–Study metric, i.e., the statistical distance. Wootters \([14]\) also investigated from a statistical viewpoint the distance between two rays, and obtained

\[
d(\psi, \varphi) = \cos^{-1} |\langle \psi | \varphi \rangle|.
\]

This is identical, up to a constant factor, to the geodesic distance as measured by the Fubini–Study metric \([2]\). Theorem 6, together with the following Theorem 7, reveals a deeper connection between them.

### 4. Parameter Estimation for Pure State Models

In this section, we give a parameter estimation theory for pure state models based on the SLD. Suppose we are given an \( n \)-parameter pure state model \((3)\). In order to handle simultaneous probability distributions of possibly mutually non-commuting observables, an extended framework of measurement theory is needed \((p. 53 [7], p. 50 [8])\). An estimator for \( \theta \) is identified to a generalized measurement which takes values on \( \Theta \). The expectation vector with respect to the measurement \( M \) at the state \( \rho_\theta \) is defined as

\[
E_\theta[M] = \int \hat{\theta} P^M_\theta (d\hat{\theta}).
\]
The measurement $M$ is called unbiased if $E_{\theta}[M] = \theta$ holds for all $\theta \in \Theta$, i.e.,
\[
\int \hat{\theta}^j P_{\theta}^M(d\hat{\theta}) = \theta^j, \quad (j = 1, \cdots, n).
\]
(12)

Differentiation yields
\[
\int \hat{\theta}^j \frac{\partial}{\partial \theta^k} P_{\theta}^M(d\hat{\theta}) = \delta^j_k, \quad (j, k = 1, \cdots, n).
\]
(13)

If (12) and (13) hold at a certain $\theta$, $M$ is called locally unbiased at $\theta$. Obviously, $M$ is unbiased iff $M$ is locally unbiased at every $\theta \in \Theta$. Letting $M$ be a locally unbiased measurement at $\theta$, we define the covariance matrix $V_{\theta}[M] = [v_{\theta}^{jk}] \in \mathbb{R}^{n \times n}$ with respect to $M$ at the state $\rho_{\theta}$ by
\[
v_{\theta}^{jk} = \int (\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k) P_{\theta}^M(d\hat{\theta}).
\]
(14)

A lower bound for $V_{\theta}[M]$ is given by the following theorem, which is a pure state version of quantum Cramér–Rao theorem.

**Theorem 7:** Given a pure state model $\rho_{\theta}$, the following inequality holds for any locally unbiased measurement $M$:
\[
V_{\theta}[M] \geq \left( J_{\theta}^S \right)^{-1}.
\]
(15)

**Proof:** It is proved in almost the same way as the strictly positive case (p. 274 [8]), except that $\langle \cdot, \cdot \rangle_{\rho_{\theta}}$ is a pre-inner product now.

When the model is one dimensional, the measurement $M$ is identified with a certain self-adjoint operator $T$, and the inequalities in the theorem become scalar, i.e.,
\[
V_{\theta}[T] \geq \frac{1}{J_{\theta}^S}.
\]
(16)

In this case, the lower bound $1/ \text{Tr} \rho_{\theta}(L_{\theta}^S)^2$ can be attained by the unbiased estimators
\[
T = \theta I + \frac{2}{J_{\theta}^S} \frac{d\rho_{\theta}}{d\theta} + K_{\theta}, \quad \forall K_{\theta} \in \mathcal{K}_{sa}(\rho_{\theta}),
\]
(17)

where $I$ is the identity. Since $d\rho_{\theta}/d\theta$ and $K_{\theta}$ do not commute in general, the measurement which attains the lower bound (16) is not determined uniquely. This fact provides significant features in the pure state estimation theory.

On the other hand, when the dimension $n \geq 2$, the matrix equality in (15) cannot be attained in general, because of the impossibility of the
exact simultaneous measurement of non-commuting observables (in von Neumann’s sense). We must, therefore, abandon the strategy of finding the measurement that minimizes the covariance matrix itself. Rather, we often adopt another strategy as follows: Given a positive definite real matrix \( G = [g_{jk}] \in \mathbb{R}^{n \times n} \), find the measurement \( M \) that minimizes the quantity

\[
\text{tr} \; GV_\theta[M] = \sum_{jk} g_{jk} v_\theta^{jk}.
\]

(18)

If there is a constant \( C \) such that \( \text{tr} \; GV_\theta[M] \geq C \) holds for all \( M \), \( C \) is called a Cramér–Rao type bound or simply a CR bound, which may depend on both \( G \) and \( \theta \). For instance, it has been shown that the following quantity is a CR bound [8, 11].

\[
C^S = \text{tr} \; G(J_\theta^S)^{-1}.
\]

This bound is, however, not always the most informative one unless \( n = 1 \). For instance, it has been shown that the CR bound based on the right logarithmic derivative is the most informative one for coherent models [3, 4]. Nevertheless, there have been few results that derived the most informative CR bounds, as the strictly positive case. The construction of the general quantum parameter estimation theory for \( n \geq 2 \) is left for future study.

5. Examples

Here we give two examples of one-parameter pure state estimation. Let us first consider a model of the form

\[
\rho_\theta = e^{i\theta H/\hbar} \rho_0 e^{-i\theta H/\hbar}.
\]

(19)

Here, \( H \) is the time independent Hamiltonian of the system, \( \hbar \) the Planck constant, and \( \theta \) the time parameter. We first assume \( \rho_0 > 0 \). According to the one-parameter estimation theory for strictly positive models [10],

\[
V[T] \geq \frac{1}{J_\theta^S} \geq \frac{1}{J(L_\theta)}
\]

(20)

holds, where \( L_\theta \) is any logarithmic derivative which satisfy (4) and \( J(L_\theta) = (L_\theta, L_\theta)_{\rho_0} \). The first equality is attained iff \( T = \theta I + L_\theta^S / J_\theta^S \), and the second equality holds iff \( L_\theta = L_\theta^S \). (Note \( L_\theta^S \) is unique for strictly positive models.)

Now, \( L_\theta^A = -2i(H - \text{Tr} \rho_0 H) / \hbar \) is the ALD for the model (19) and the corresponding Cramér–Rao inequality becomes

\[
V_\theta[T] \geq \frac{\hbar^2}{4V_\theta[H]}.
\]

(21)
where \( T \) is an arbitrary unbiased estimator \( T \) for the time parameter \( \theta \). This inequality is nothing but a time-energy uncertainty relation. It is notable that the lower bound cannot be attained for any \( T \) since \( L^A_\theta \) is not the SLD for strictly positive models. We next assume that \( \rho_0 \) is pure. In this case, Theorem 6 and (16) (17) assert that the equality in (21) is locally attainable. This is a significant difference between the strictly positive models and the pure state models. Since both the ALD \( L^A_\theta = -2i(H - \text{Tr} \rho_0 H) / \hbar \) and the SLD–Fisher information for the pure state models \( J^S_\theta = 2 \text{Tr}(d\rho_0 / d\theta)^2 \) can be obtained directly from the Liouville–von Neumann equation, this result is not specific to the case where the Hamiltonian is time independent, but is quite general.

We next consider the efficiency of an estimator. An unbiased estimator \( T \) is called efficient if the equality in (16) holds for all \( \theta \in \Theta \). It has been proved in [10] that a one-parameter model \( \rho_\theta \) has an efficient estimator iff the model takes the form
\[
\rho_\theta = e^{\frac{i}{2} \beta(\theta) T - \gamma(\theta)} \rho_0 e^{\frac{i}{2} \beta(\theta) T - \gamma(\theta)},
\]
(22)
where \( \beta(\theta), \gamma(\theta) \) are real functions. Now, let us consider a model of the form
\[
\rho_\theta = e^{if(\theta)A} \rho_0 e^{-if(\theta)A},
\]
where \( f(\theta) \) is a real monotonic odd function and \( A \in \mathcal{L}_{sa} \). If \( \rho_0 > 0 \), then it is shown that there exists an efficient estimator for \( \theta \) only when \( A \) is a canonical observable [5]. This cannot be the case unless the underlying Hilbert space \( \mathcal{H} \) is infinite dimensional. On the other hand, if \( \rho_0 \) is pure, then there may exist an efficient estimator even if \( A \) is not canonical, because of the uncertainty \( K_\theta \in \mathcal{K}_{sa}(\rho_\theta) \) in (17). For instance, the spin 1/2 model
\[
f(\theta) = \frac{1}{2} \left( \frac{\pi}{2} - \cos^{-1} \theta \right), \quad A = \sigma_y, \quad \rho_0 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},
\]
has an efficient estimator \( \sigma_z \) for the parameter \( \theta \). Indeed, this model permits another form
\[
\rho_\theta = \sqrt{1 - \theta^2} \exp \left[ \frac{1}{4} \log \frac{1 + \theta}{1 - \theta} \sigma_z \right] \rho_0 \exp \left[ \frac{1}{4} \log \frac{1 + \theta}{1 - \theta} \sigma_z \right].
\]
(23)
This is not a paradox since, in the pure state model, the estimator which attains the Cramér–Rao bound (16) is adjustable for every points \( \rho_\theta \) up to the uncertainty of the kernel \( \mathcal{K}_{sa}(\rho_\theta) \), see also [5].
6. Conclusions

A quantum estimation theory for pure state models was presented. We first derived some mathematical lemmas to treat the pure state models and derived quantum counterpart of the Fisher metric. The statistical significance of the Fubini–Study metric was also stressed. We then formulated the one-parameter pure state estimation theory based on the symmetric logarithmic derivative and disclosed the characteristics of pure states. Some examples were given in order to clarify the difference between the pure state models and the strictly positive models. The construction of the general quantum multi-parameter estimation theory is, however, left for future study, as the strictly positive model case.

References

CHAPTER 18

Geometry of Quantum Estimation Theory

Akio Fujiwara*

1. Introduction

Quantum estimation theory, the quantum counterpart of the classical statistical estimation theory, originated in the study of finding the optimal detection of continuous signals in optical communication theory. The purpose of quantum estimation theory is to find the optimal measurement that gives the best estimates of the true state which lies in a certain parametric family of quantum states. A quantum statistical model is a family of density operators \( \rho_\theta \) defined on a certain separable Hilbert space \( \mathcal{H} \) with finite-dimensional real parameters \( \theta = (\theta^i)_{i=1}^n \) which are to be estimated statistically. It was Helstrom [14] who founded the quantum estimation theory and successfully introduced the symmetric logarithmic derivative for the one-parameter estimation theory as a quantum counterpart of the logarithmic derivative in the classical estimation theory. Yuen and Lax [32] investigated the two-parameter case in the expectation parameter estimation for quantum Gaussian models by introducing the right logarithmic derivative, where each parameter corresponds to mutually non-commuting two observables. These results have provided theoretical background of optical communication theory. Though many other researches have been made in multi-parameter estimation theory, the general solution is far from reaching as yet.

For the reader’s convenience, we first give a brief summary of the conventional quantum parameter estimation theory. For details, consult [6, 15, 17].

Let \( \mathcal{H} \) be a separable Hilbert space which corresponds to a physical system with inner product \( \langle \phi|\psi \rangle \), \( \langle \phi, \psi \in \mathcal{H} \rangle \), and \( \mathcal{L} \) and \( \mathcal{L}_{sa} \) be, respectively,
the set of all (bounded) linear operators and all self-adjoint operators on \( \mathcal{H} \). A quantum state is represented by a density operator \( \rho \in \mathcal{L}_{sa} \) which satisfies \( \rho \geq 0 \) and \( \text{Tr } \rho = 1 \). A state \( \rho \) is called pure if \( \text{rank } \rho = 1 \). In order to handle simultaneous probability distributions of possibly mutually non-commuting observables, an extended framework of measurement theory is needed (p. 53 [15], p. 50 [17]). A generalized measurement \( \{ M(B) \}_{B \in \mathcal{F}} \) on a measurable space \( (\Omega, \mathcal{F}) \) is an operator-valued set function which satisfy the following axioms:

1. \( M(\phi) = 0, \quad M(\Omega) = I \),
2. \( M(B) = M(B)^\ast \geq 0, \quad (\forall B \in \mathcal{F}) \),
3. \( M(\bigcup_j B_j) = \sum_j M(B_j), \quad (\text{for all at most countable disjoint sequence } \{B_j \} \subset \mathcal{F}) \).

By fixing a state \( \rho \) and a measurement \( M \), outcomes of the measurement form random variables, whose simultaneous distribution is given by \( P^M_\rho(B) = \text{Tr } \rho M(B) \). In particular, a measurement \( M \) is called simple if it satisfies, in addition to the above three axioms, \( M(B)^2 = M(B), \quad (\forall B \in \mathcal{F}) \), or equivalently \( M(B_1)M(B_2) = 0, \quad (\forall B_1 \cap B_2 = \phi) \). In the following, simple measurement is denoted by \( E \).

Here we give an example of generalized measurement, which is called the random measurement (p. 70 [15]). Imagine \( N \) measuring apparatuses, which yield outcomes from the same finite set whose elements being labeled with the integers \( j \). Suppose the \( n \)th apparatus is described by a simple measurement \( \{ E^{(n)}_j \}_j \). Select an apparatus \( n \) \((n = 1, 2, \ldots, N)\) with probability \( \xi_n \) and apply to the physical system. Then the probability for obtaining the \( j \)th outcome is

\[
P^M_\rho(j) = \sum_{n=1}^{N} \xi_n \text{Tr } \rho E^{(n)}_j = \text{Tr } \rho M_j,
\]

where

\[
M_j = \sum_{n=1}^{N} \xi_n E^{(n)}_j
\]

form a generally non-orthogonal measurement.

Let

\[
\mathcal{S} = \{ \rho_\theta : \rho_\theta = \rho_\theta^\ast > 0, \quad \text{Tr } \rho_\theta = 1, \quad \theta \in \Theta \subset \mathbb{R}^n \}
\]

be the statistical parametric model composed of strictly positive density operators. Here, \( \theta \) is the parameter to be estimated statistically. An estimator for \( \theta \) is identified to a generalized measurement which takes values
on $\Theta$. The expectation vector with respect to the measurement $M$ at the state $\rho_{\theta}$ is defined as

$$E_{\theta}[M] = \int \hat{\theta} P_{\theta}^M(d\hat{\theta}).$$

The measurement $M$ is called *unbiased* if $E_{\theta}[M] = \theta$ holds for all $\theta \in \Theta$, i.e.,

$$\int \hat{\theta}^j P_{\theta}^M(d\hat{\theta}) = \theta^j, \quad (j = 1, \cdots, n). \tag{2}$$

Differentiation yields

$$\int \hat{\theta}^j \frac{\partial}{\partial \theta^k} P_{\theta}^M(d\hat{\theta}) = \delta^j_k, \quad (j, k = 1, \cdots, n). \tag{3}$$

If (2) and (3) hold at a certain $\theta$, $M$ is called *locally unbiased* at $\theta$. Obviously, $M$ is unbiased iff $M$ is locally unbiased at every $\theta \in \Theta$. A locally unbiased measurement is accompanied by $n$ self-adjoint operators:

$$X^j = \int (\hat{\theta}^j - \theta^j) M(d\hat{\theta}), \quad (j = 1, \cdots, n). \tag{4}$$

The locally unbiasedness conditions (2) (3) are then re-expressed as

$$\text{Tr} \rho_{\theta} X^j = 0, \quad \text{Tr} \frac{\partial \rho_{\theta}}{\partial \theta^k} X^j = \delta^j_k, \quad (j, k = 1, \cdots, n). \tag{5}$$

Since these self-adjoint operators $\{X^j\}$ are usefully employed in the following, we may call them *locally unbiased operators*. Letting $M$ be a locally unbiased measurement at $\theta$, we define the covariance matrix $V_{\theta}[M] = [v_{\theta}^{jk}] \in \mathbb{R}^{n \times n}$ with respect to $M$ at the state $\rho_{\theta}$ by

$$v_{\theta}^{jk} = \int (\hat{\theta}^j - \theta^j)(\hat{\theta}^k - \theta^k) P_{\theta}^M(d\hat{\theta}). \tag{6}$$

In order to obtain lower bounds for $V_{\theta}[M]$, let us consider a quantum analogue of the logarithmic derivative denoted by $L_{\theta}$:

$$\frac{\partial \rho_{\theta}}{\partial \theta^j} = \frac{1}{2}[\rho_{\theta} L_{\theta,j} + L_{\theta,j}^* \rho_{\theta}]. \tag{7}$$

For instance,

$$\frac{\partial \rho_{\theta}}{\partial \theta^j} = \frac{1}{2}[\rho_{\theta} L_{\theta,j}^S + L_{\theta,j}^S \rho_{\theta}], \quad L_{\theta,j}^S = L_{\theta,j}^{S*} \quad \tag{8}$$

defines the *symmetric logarithmic derivative* (SLD) $L_{\theta,j}^S$ introduced by Helstrom [14], and

$$\frac{\partial \rho_{\theta}}{\partial \theta^j} = \rho_{\theta} L_{\theta,j}^R \quad \tag{9}$$
defines the right logarithmic derivative (RLD) $L^R_{\theta,j}$ introduced by Yuen and Lax [32]. Thus, (7) defines a certain class of logarithmic derivatives. Correspondingly, we define the quantum analogue of Fisher information matrix $J_{\theta} = [(L_{\theta,j}, L_{\theta,k})_{\rho\theta}]$, where the inner product $(\cdot, \cdot)_{\rho}$ on $L$ is defined by

$$(A, B)_{\rho} = \text{Tr} \rho BA^*.$$ 

(10)

We also define another inner product on $L$ as

$$\langle A, B \rangle_{\rho} = \frac{1}{2} \text{Tr} \rho (BA^* + A^*B).$$

(11)

Then, the following quantum version of Cramér–Rao theorem holds.

**Proposition 1:** For any locally unbiased measurement $M$, the following inequality holds:

$$V_{\theta}[M] \geq (\text{Re} J_{\theta})^{-1},$$

(12)

where $\text{Re} J_{\theta} = (J_{\theta} + \overline{J_{\theta}})/2$. In particular, for the SLD, $J^S_{\theta} = \text{Re} J_{\theta} = \left[ \langle L^S_{\theta,j}, L^S_{\theta,k} \rangle_{\rho\theta} \right]$ is called the SLD–Fisher information matrix. Moreover, for the RLD,

$$V_{\theta}[M] \geq (J^R_{\theta})^{-1}$$

(13)

holds, where $J^R_{\theta} = \left[ \langle L^R_{\theta,j}, L^R_{\theta,k} \rangle_{\rho\theta} \right]$ is called the RLD–Fisher information matrix.

When the model is one dimensional, the inequalities in the theorem become scalar. In this case, it is shown that the lower bound $(\text{Re} J_{\theta})^{-1} = (J_{\theta})^{-1}$ becomes most informative, i.e., it takes the maximal value, iff the SLD is adopted, and the corresponding lower bound $(J^S_{\theta})^{-1} = 1/\text{Tr} \rho_{\theta}(L^S_{\theta})^2$ can be attained by the estimator $T = \theta I + L^S_{\theta}/J^S_{\theta}$, where $I$ is the identity. Thus, the one parameter quantum estimation theory is quite analogous to the classical one when the SLD is used.

On the other hand, for the dimension $n \geq 2$, the matrix equalities in (12) and (13) cannot be attained in general because of the impossibility of the exact simultaneous measurement of non-commuting observables. In other words, this inequality is not so useful in the quantum multi-parameter estimation theory, which is in contrast to the classical estimation theory. We must, therefore, abandon the strategy of finding the measurement that minimizes the covariance matrix itself. Rather, we often adopt
an alternative strategy as follows: Given a non-negative definite real matrix 
\( G = [g_{jk}] \in \mathbb{R}^{n \times n} \), find the measurement \( M \) that minimizes the quantity
\[
\text{tr} \ G V_\theta [M] = \sum_{jk} g_{jk} v_{\theta}^{jk}.
\] (14)

If there is a constant \( C \) such that \( \text{tr} \ G V_\theta [M] \geq C \) holds for all \( M \), \( C \) is called a Cramér–Rao type bound or simply a CR bound, which may depend on both \( G \) and \( \theta \). For instance, it is shown that the following two quantities are both CR bounds \([23]\).
\[
C^S = \text{tr} \ G (J^S_\theta)^{-1},
\] (15)
\[
C^R = \text{tr} \ G \text{Re} (J^R_\theta)^{-1} + \text{tr} \text{abs} G \text{Im} (J^R_\theta)^{-1}.
\] (16)

Here, for a given matrix \( X \), \( \text{Im} X = (X - \overline{X})/2i \), and \( \text{tr} \text{abs} X \) denotes the absolute sum of the eigenvalues of \( X \). Let us call these CR bounds, respectively, the SLD–bound and the RLD–bound. The most informative CR bound is the maximum value of such \( C \) for given \( G \) and \( \theta \). Yuen and Lax \([32]\) proved that the above \( C^R \) is most informative for the Gaussian model, and they explicitly constructed the optimum measurement which attains \( C^R \). Holevo (p. 285 \([17]\)) derived another CR bound which, though an implicit form, is not less informative than \( C^S \) and \( C^R \). Nagaoka \([23]\) (cf. \([24]\)) investigated in detail the relation between these CR bounds. He also derived a new CR bound for two-dimensional models, which is not less informative than Holevo’s one, and obtained explicitly the most informative CR bound specific to the spin 1/2 model. The construction of the general quantum parameter estimation theory for \( n \geq 2 \) is left for future study.

2. Dualistic Geometry Based on the Logarithmic Derivatives

As was mentioned in Part I\(^1\), there exist a variety of manners to define quantum counterparts of geometrical notions used in the classical estimation theory such as divergence, logarithmic derivative, Fisher information, Cramér–Rao inequality, etc. Moreover, only a few of them have been proved crucial in solving concrete quantum estimation problems so far \([14, 23, 32]\). Therefore, if we invoke an easy analogy of a certain aspect of classical information geometry, we cannot expect fruitful geometrical viewpoints. In order not to construct abundance of useless imitations, we must stand upon

a deeper understanding of inherent geometrical aspects of quantum estimation theory. In this section, we offer a new geometrical structure of quantum states and investigate its significance in quantum estimation theory [10].

In Section 2.1, we define a dualistic geometrical structure of quantum states based on the symmetric logarithmic derivatives. This structure has a non-vanishing torsion field in general, so that there exists no divergence function on the quantum space. In Section 2.2, the autoparallelity of a quantum state model is investigated in detail, which also clarifies the difference between classical and various quantum criterions of estimation. In Section 2.3, a condition for a one-parameter unitary model to be autoparallel, i.e., to have an efficient estimator, is derived. This condition indicates the importance of canonical observables in an estimation theoretical viewpoint. Some examples are also presented.

2.1. Dualistic structure on quantum states

For brevity, the readers are assumed familiar with some elementary knowledge of classical information geometry [2]. Let $\mathcal{H}$ be a separable Hilbert space which corresponds to a physical system with inner product $\langle \phi | \psi \rangle$, $(\phi, \psi \in \mathcal{H})$. A quantum state is represented by a density operator $\rho$ which satisfies $\rho = \rho^* \geq 0$ and $\text{Tr} \rho = 1$. In this paper, we further assume $\rho > 0$ for simplicity. Denote by $\mathcal{P}$ the totality of such strictly positive density operators. Let us start with definitions of quantum analogues of some basic notions in classical theory. We first define the mixture representation of a tangent vector $\partial$ at $\rho$ by the isomorphism $\partial \leftrightarrow \partial \rho$ and call $\partial \rho$ the $m$-tangent vector. Denote their totality by $T^{(m)}(\mathcal{P})$. On the other hand, as was mentioned in the previous section, there exist a variety of candidates of the quantum counterpart of the exponential representation of the tangent vector due to the non-commutativity. Nevertheless, we here adopt the isomorphism $\partial \leftrightarrow L$ defined by

$$\partial \rho = \frac{1}{2} [\rho L + L \rho], \quad L = L^*$$

as the exponential representation, since it gives the definition of the so-called symmetric logarithmic derivative (SLD) which leads us to the most informative Cramér–Rao bound in the one parameter estimation theory [22]. Let us call $L$ the e-tangent vector and denote their totality by $T^{(e)}(\mathcal{P})$. These isomorphisms are characterized by the following schemata:
• m-tangent vector
\[
T_{\rho}(\mathcal{P}) \cong T_{\rho}^{(m)}(\mathcal{P}) = \{ G ; \ G = G^*, \ \text{Tr} \ G = 0 \}
\]
\[
\partial \cong G \quad \text{where} \quad \partial \rho = G
\]
• e-tangent vector
\[
T_{\rho}(\mathcal{P}) \cong T_{\rho}^{(e)}(\mathcal{P}) = \{ H ; \ H = H^*, \ \text{Tr} \ \rho H = 0 \}
\]
\[
\partial \cong H \quad \text{where} \quad \partial \rho = \frac{1}{2} [\rho H + H \rho], \ H = H^*
\]

Correspondingly, we define two kinds of parallel translations:

• m-parallel translation
\[
T_{\rho}(\mathcal{P}) \cong T_{\rho}^{(m)}(\mathcal{P}) \ni G
\]
\[
\downarrow \quad \downarrow \quad \downarrow
\]
\[
T_{\rho'}(\mathcal{P}) \cong T_{\rho'}^{(m)}(\mathcal{P}) \ni G
\]
• e-parallel translation
\[
T_{\rho}(\mathcal{P}) \cong T_{\rho}^{(e)}(\mathcal{P}) \ni H
\]
\[
\downarrow \quad \downarrow \quad \downarrow
\]
\[
T_{\rho'}(\mathcal{P}) \cong T_{\rho'}^{(e)}(\mathcal{P}) \ni H - \text{Tr} \ \rho' H
\]

These parallel translations naturally define two affine connections \( \nabla^{(m)} \) and \( \nabla^{(e)} \), which are called the mixture connection (m-connection) and the exponential connection (e-connection). Further, by using the symmetrized inner product \( \langle A, B \rangle_{\rho} = \frac{1}{2} \text{Tr} \rho (A^* B + BA^*) \), we define the Riemannian metric at \( \rho \) by
\[
g_{\rho}(\partial_i, \partial_j) = \langle L_i, L_j \rangle_{\rho} = \frac{1}{2} \text{Tr} \ \rho (L_i L_j + L_j L_i).
\]

Since the metric can be written as \( g_{\rho}(\partial_i, \partial_j) = \text{Tr} (\partial_i \rho) L_j \), the two connections \( \nabla^{(m)} \) and \( \nabla^{(e)} \) are mutually dual with respect to the metric \( g \) in the following sense: For arbitrary vector fields \( X, Y, Z \) on \( \mathcal{P} \),
\[
X g(Y, Z) = g(\nabla^{(m)}_X Y, Z) + g(Y, \nabla^{(e)}_X Z)
\]
holds. It is evident that both curvature tensors with respect to the connections \( \nabla^{(m)} \) and \( \nabla^{(e)} \) vanish, since the two parallel translations are defined independently of the choice of the path connecting \( \rho \) and \( \rho' \). The torsion tensor with respect to the m-connection also vanishes, whereas the torsion
tensor $T^{(e)}$ with respect to the e-connection does not vanish in general since it becomes
\[ T^{(e)}(\partial_j, \partial_k)\rho = \frac{1}{4}[[L_j, L_k], \rho]. \quad (17) \]
From this fact, the divergence function does not exist on $\mathcal{P}$ in general.

Let us prove (17). We first construct two e-parallel vector fields $X_1$ and $X_2$ by translating two arbitrarily fixed tangent vectors $\partial_1, \partial_2$ at $\rho_0$ with respect to the e-connection. Then
\[ \nabla^{(e)}_{X_1} X_2 = \nabla^{(e)}_{X_2} X_1 = 0, \]
and the torsion becomes
\[ T^{(e)}(X_1, X_2) = \nabla^{(e)}_{X_1} X_2 - \nabla^{(e)}_{X_2} X_1 - [X_1, X_2] = -[X_1, X_2]. \]
Letting the e-representation of the tangent vector $\partial_j$ ($j = 1, 2$) at $\rho_0$ be $L_j$, the e-representation of the tangent vector $(X_j)_\rho$ becomes
\[ (X_j)_\rho \simeq (X^{(e)}_j)_\rho = L_j - \text{Tr} \rho L_j, \]
which acts on $\rho$ as
\[ X_j \rho = \frac{1}{2} \left[ \rho (X^{(e)}_j)_\rho + (X^{(e)}_j)_\rho \rho \right] = \frac{1}{2} \left[ (\rho L_j + L_j \rho) - 2 \rho \text{Tr} \rho L_j \right]. \quad (18) \]
Since the quantity $X_2 X_1 \rho$ describe the change of $X_1 \rho$ when $\rho$ is slightly moved along $X_2$-direction,
\[ X_2 X_1 \rho = \frac{1}{2} \left[ (X^{(e)}_2 \rho)(X^{(e)}_1 \rho) + (X^{(e)}_1 \rho)(X^{(e)}_2 \rho) \right] \]
\[ = \frac{1}{2} \left[ ((X_2 \rho)L_1 + L_1 (X_2 \rho)) - 2 (X_2 \rho) \text{Tr} \rho L_1 - 2 \rho \text{Tr} (X_2 \rho) L_1 \right]. \]
Substituting (18) into the above equation, we have
\[ X_2 X_1 \rho = \frac{1}{4} (\rho L_2 L_1 + L_1 L_2 \rho) + \{ \text{symmetric terms with respect to } L_1, L_2 \}. \]
We can evaluate $X_1 X_2 \rho$ in the same way, yielding
\[ T^{(e)}(X_1, X_2) \rho = [X_1, X_2] \rho = \frac{1}{4}[[L_1, L_2], \rho]. \]
Now, since $\rho_0$ is arbitrary, the torsion at any point $\rho \in \mathcal{P}$ becomes
\[ T^{(e)}(\partial_j, \partial_k)\rho = \frac{1}{4}[[L_j, L_k], \rho], \quad (19) \]
where $L_j$ is the e-representation of the tangent vector $\partial_j$ at $\rho$. For an arbitrary submanifold $M$, the torsion $T^{(e)}_M$ is obtained by projecting $T^{(e)}$ onto the tangent space of $M$ with respect to the Riemannian metric.
2.2. Autoparallelity in quantum estimation theory

In classical estimation theory, one of the most important geometrical notion is the autoparallelity of a model with respect to the e-connection, which is an equivalent condition for the existence of the efficient estimator of the model [2]. An e-autoparallel model is also called an exponential family which takes the form

\[ p_\theta(x) = \exp \left[ \theta^i f_i(x) - \psi(\theta) \right], \]

where \( \theta = (\theta^1, \cdots, \theta^n) \) is the \( n \)-dimensional e-affine parameter to be estimated statistically, \( \psi(\theta) \) the normalization factor, and Einstein’s summation convention \( \theta^i f_i(x) = \sum_i \theta^i f_i(x) \) is used. In this section, we investigate the quantum counterpart of this notion in detail.

Let us first consider some conditions relevant to the efficiency of the estimator. We add some terminologies:

1 A locally unbiased measurement \( M \) is called locally efficient at \( \theta \) if

\[ V_\theta[M] \leq V_\theta[M'] \]

holds for every locally unbiased measurement \( M' \) at \( \theta \). A measurement \( M \) is called efficient if \( M \) is locally efficient for all \( \theta \).

2 Given an arbitrary weight (real symmetric non-negative matrix) \( G \), a locally unbiased measurement \( M \) is called \( G \)-locally efficient at \( \theta \) if

\[ \text{tr} \, G V_\theta[M] \leq \text{tr} \, G V_\theta[M'] \]

holds for every locally unbiased measurement \( M' \) at \( \theta \). Given an arbitrary weight field \( G = \{ G_\theta | \, \theta \in \Theta \} \), a measurement \( M \) is called \( G \)-efficient if \( M \) is \( G_\theta \)-locally efficient for all \( \theta \). In particular, if \( G_\theta \equiv G \) for all \( \theta \), \( M \) is called \( G \)-efficient.

There exist some evident relations between these notions, which are listed in the following propositions to put the issues in order.

**Proposition 2:** The following conditions for a measurement \( M \) are equivalent:

(i) \( M \) is locally efficient at \( \theta \).
(ii) For all \( G > 0 \), \( M \) is \( G \)-locally efficient at \( \theta \).
(iii) For all \( v = (v_1, \cdots, v_n) \in \mathbb{R}^n \), \( M \) is \( v^T v \)-locally efficient at \( \theta \).
(iv) \( M \) is locally unbiased at \( \theta \) and \( V_\theta[M] = (J_\theta)^{-1} \) holds at \( \theta \).

**Proposition 3:** The following conditions for a measurement \( M \) are equivalent:
(i) $M$ is efficient.
(ii) For all $G$, $M$ is $G$–efficient.
(iii) For all $G > 0$, $M$ is $G$–efficient.
(iv) For all $v = (v_1, \cdots, v_n) \in \mathbb{R}^n$, $M$ is $v^T v$–efficient.
(v) $M$ is unbiased and $V_\theta[M] = (J_\theta)^{-1}$ holds for all $\theta$.

Proposition 4: Given a model $S = \{\rho_\theta \mid \theta \in \Theta\}$, consider the following conditions:

(i) $S$ has an efficient measurement.
(ii) $S$ has (possibly $G$-dependent) $G$–efficient measurements $M_G$ for all $G$.
(iii) $S$ has (possibly $G$-dependent) $G$–efficient measurements $M_G$ for all $G > 0$.
(iv) $S$ has (possibly $v^T v$-dependent) $v^T v$–efficient measurements $M_{\{v^T v\}}$ for all $v = (v_1, \cdots, v_n) \in \mathbb{R}^n$.
(v) There exists a certain $G$ for which $S$ has a $G$–efficient measurement.
(vi) $S$ is $c$-autoparallel, all the SLDs commute, and $\theta$ is an $m$-affine coordinate system.

In classical theory, all these conditions are equivalent. In quantum case, however, only the following relations hold in general:

$$
(i) \iff (vi), \quad (i) \implies (ii) \implies \left\{ \begin{array}{l}
(iii) \\
(iv) \\
(v)
\end{array} \right.
$$

Proposition 4 clarifies the difference between the classical and the quantum estimation theories. Indeed, since the SLDs do not commute in general, we cannot expect the existence of an efficient estimator. We therefore have adopted another strategy of minimizing the weighted sum of the covariances (14) instead of the classical minimization strategy of the covariance matrix itself. Nevertheless, Proposition 4 also indicates that there exist a variety of strategies which are not mutually equivalent in general. Indeed, the relations between the conditions (iii) (iv) (v) in Proposition 4 are not yet clear.

By the way, the condition (iv) in Proposition 4 is closely related to the one-parameter quantum estimation theory which is rather well-established so far. Indeed, the restriction of the Cramér–Rao inequality (12) in the $v$-direction

$$
v V_\theta[M] v^T \geq v (J_\theta)^{-1} v^T
$$

(20)
gives explicitly the locally attainable lower bound for $\text{tr} (v^T v) V_\theta[M]$. More precisely, the following lemma holds.
Lemma 5: Given a model \( S = \{ \rho_\theta \mid \theta = (\theta^1, \ldots, \theta^n) \in \Theta \} \) and an arbitrary \( v \in \mathbb{R}^n \), consider the differential equation
\[
\frac{d\theta^i}{dt} = v_j J^{ji},
\]
(21)
where \( J^{ij} \) is the \((i, j)\) entry of \((J_\theta)^{-1}\), and denote by \( \theta(t) \) the solution of (21) for an arbitrarily fixed initial condition \( \theta(0) \in \Theta \). Then the restricted Cramér–Rao inequality (20) can be regarded as the one dimensional Cramér–Rao inequality for the one parameter sub-model \( \rho_{\theta(t)} \).

Proof: Let \( L_j \) be the SLD for the parameter \( \theta^j \). For every locally unbiased measurement \( M \) of \( \theta \),
\[
T(v) = \int v_i \hat{\theta}^i M(d\hat{\theta})
\]
becomes a locally unbiased estimator of the parameter \( \theta(v) = v_i \theta^i \) at \( \theta \) and satisfies
\[
\langle T(v) - \theta(v), L_j \rangle_{\rho_\theta} = \text{Re} \text{Tr} \rho_\theta \int v_i \hat{\theta}^i M(d\hat{\theta}) L_j = v_i \text{Re} \text{Tr} L_j \rho_\theta \int \hat{\theta}^i M(d\hat{\theta}) = v_i \frac{\partial}{\partial \theta^j} \rho_\theta \int \hat{\theta}^i M(d\hat{\theta}) = v_i \frac{\partial}{\partial \theta^j} \theta^i = v_j.
\]
Then the orthogonal projection of \( T(v) - \theta(v) \) onto the SLD–tangent space span \( \{L_j\}_{j=1}^n \) with respect to the inner product \( \langle \cdot, \cdot \rangle_{\rho_\theta} \) is \( S(v) = v_j J^{ji} L_i \) since, by expressing \( S(v) = c^i L_i \),
\[
v_j = \langle T(v) - \theta(v), L_j \rangle_{\rho_\theta} = \langle S(v), L_j \rangle_{\rho_\theta} = c^i \langle L_i, L_j \rangle_{\rho_\theta} = c^i J_{ij},
\]
where \( J_{ij}, J^{ij} \) are the \((i, j)\) entries of \( J_\theta, (J_\theta)^{-1} \), respectively. Therefore, the inequality
\[
\langle T(v) - \theta(v), T(v) - \theta(v) \rangle_{\rho_\theta} \geq \langle S(v), S(v) \rangle_{\rho_\theta}
\]
characteristic of the projection is nothing but (20). Now, let us determine the sub-model \( \rho_{\theta(t)} \) whose Cramér–Rao inequality becomes (20), i.e., whose SLD is \( S(v) \). Observing
\[
\frac{1}{2} [S(v)\rho_{\theta(t)} + \rho_{\theta(t)} S(v)] = \frac{d\rho_{\theta(t)}}{dt} = \frac{\partial \rho_{\theta(t)}}{\partial \theta^i} \frac{d\theta^i}{dt} = \frac{1}{2} [L_i \rho_{\theta(t)} + \rho_{\theta(t)} L_i] \frac{d\theta^i}{dt},
\]
we have
\[
S(v) = L_i \frac{d\theta^i}{dt}.
\]
Therefore, the desired sub-model is determined by the following differential equation

$$\frac{d\theta^i}{dt} = v_j J^{ji}, \quad \theta(0) = \theta_0.$$ 

Since

$$\frac{d}{dt} [v_i \theta^i] = v_i J^{ij} v_j > 0,$$

we can take \( t = v_i \theta^i \), which proves the lemma.

This lemma immediately leads us to the following theorem, which shows that the condition (iv) in Proposition 4 is a sufficient condition for the model to be totally e-geodesic.

**Theorem 6:** If a model \( S = \{ \rho_{\theta}; \theta = (\theta^1, \ldots, \theta^n) \in \Theta \} \) has (possibly \( v \)-dependent) \( \mathbf{v}^T \mathbf{v} \)-efficient measurements for all \( \mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n \), then \( S \) is totally e-geodesic and \( \theta \) is the \( m \)-affine coordinate system.

**Proof:** From the assumption, the restricted Cramér–Rao bound (20) can be attained by a certain \( \theta \)-independent measurement. Then, by invoking Lemma 5, the sub-model \( \rho_{\theta(t)} \) which is determined by \( \mathbf{v} \) and \( \rho_0 \) has an efficient estimator for \( t \). Then \( \rho_{\theta(t)} \) must be an e-geodesic and \( t \) is an \( m \)-affine parameter. Since \( \mathbf{v} \) is arbitrary, the model is totally e-geodesic and \( \theta \) is \( m \)-affine.

Theorem 6 asserts that the minimization strategy of \( \text{tr} \, G V_{\theta} [M] \), \( (G \geq 0) \) is closely related to the geometric structure of the model. Indeed, if the model is not totally e-geodesic (hence not e-autoparallel), we may not expect the existence of \( G \)-efficient measurement. Therefore the geodesic nature of a model is an important notion in quantum estimation theory.

In general, an autoparallel submanifold is automatically a totally geodesic submanifold. Conversely, a totally geodesic submanifold becomes an autoparallel submanifold if the enveloping manifold is torsion free (II, p. 53 [20]). In this sense, the assumption of Theorem 6 may have little to do with the condition for a model to be e-autoparallel. Though a necessary and sufficient condition for a model to be e-autoparallel has been obtained by Nagaoka [27], the relation between the assumption of Theorem 6 and the e-autoparallelity of the model is not yet clarified so far.
2.3. Autoparallelity in unitary models

Since a model $\rho_\theta$ is represented, in general, by a spectral decomposition

$$\rho_\theta = \int p_\theta(x) E_\theta(dx),$$

the parameter change is composed of two parts: eigenvalue part (classical part) and unitary part (purely quantum part). One extreme case is a model where $E_\theta$ is $\theta$-independent (classical statistical model). Let us call another extreme a unitary model where $p_\theta$ is $\theta$-independent. Owing to group theoretical symmetry of the physical system, we often encounter such unitary models. Therefore, it is expected that the study of unitary models may bring us some important suggestions toward the construction of general quantum estimation theory. In this section, we investigate the geodesic nature of unitary models.

Since Theorem 6 indicates the importance of the decomposition of the model into foliation of one-dimensional submanifolds, we consider here one-dimensional unitary models and derive a necessary and sufficient condition for the model to be an e-geodesic. Let us consider a one parameter unitary model of the form

$$\rho_t = e^{if(t)A} \rho_0 e^{-if(t)A},$$

(22)

where $A$ is a self-adjoint generator, $\rho_0$ a strictly positive initial density operator, $f(t)$ a real smooth monotonic odd function with one dimensional real parameter $t$. Let us recall the commutation operator $\mathcal{D}$ introduced by Holevo [17]. For an arbitrary state $\rho$ and a self-adjoint operator $X$, $\mathcal{D}$ is defined by

$$i(X \rho - \rho X) = \frac{1}{2}((\mathcal{D}X) \rho + \rho(\mathcal{D}X)), \quad (\mathcal{D}X)^* = \mathcal{D}X.$$  

(23)

This is an anti-symmetric super-operator such that $\langle A, \mathcal{D} B \rangle_\rho = -\langle \mathcal{D} A, B \rangle_\rho$ holds for all $A, B \in \mathcal{L}_{sa}$. Note that $\mathcal{D}X$ is uniquely determined iff $\rho > 0$.

**Lemma 7:** Denoting by $\mathcal{D}_t$ the commutation operator with respect to the unitary model $\rho_t$, then

$$\mathcal{D}_t A = e^{if(t)A} (\mathcal{D}_0 A) e^{-if(t)A},$$

holds, and the SLD becomes

$$L_t = f'(t) e^{if(t)A} (\mathcal{D}_0 A) e^{-if(t)A}.$$
Proof:

\[
\frac{d\rho_t}{dt} = if'(t)[A\rho_t - \rho_tA] = if'(t)e^{if(t)A}[A\rho_0 - \rho_0A]e^{-if(t)A}
\]

\[
= f'(t)e^{if(t)A} \frac{1}{2}[\rho_0(\mathfrak{D}_0A) + (\mathfrak{D}_0A)\rho_0]e^{-if(t)A}
\]

\[
= f'(t)\frac{1}{2}\left[\rho_t e^{if(t)A}(\mathfrak{D}_0A)e^{-if(t)A} + e^{if(t)A}(\mathfrak{D}_0A)e^{-if(t)A}\rho_t\right].
\]

This proves the lemma.

\[\square\]

**Lemma 8:** The unitary model (22) becomes an e-geodesic iff there exists a real function \(J(t)\) such that the self-adjoint operator

\[T = \frac{f'(0)}{J(0)} \mathfrak{D}_0A\]

satisfies the following relation

\[e^{if(t)A}Te^{-if(t)A} = \frac{f'(0)J(t)}{f'(t)J(0)}(T - t).\]

In this case, \(T\) becomes the efficient estimator of the parameter \(t\), and \(J(t)\) becomes the SLD–Fisher information of the model.

**Proof:** In general, a one-parameter model \(\rho_t\) has an efficient estimator \(T\) of the parameter \(t\) iff there exists a real function \(J(t)\) such that

\[L_t = J(t)(T - t)\]

holds. Comparing this and Lemma 7, we have

\[f'(t)e^{if(t)A}(\mathfrak{D}_0A)e^{-if(t)A} = J(t)(T - t).\]

Setting \(t = 0\), we have the definition of \(T\), which immediately leads to the lemma.

\[\square\]

**Lemma 9:** Suppose we are given self-adjoint operators \(A, B\), and real functions \(f(t), g(t)\) such that \(f(-t) = -f(t), \ g(0) = 1\). The equality of the form

\[e^{if(t)A}Be^{-if(t)A} = g(t)(B - t)\]

holds iff there exists a non-zero real number \(\mu\) such that

\[f(t) = \frac{t}{\mu}, \quad g(t) = 1, \quad [A, B] = i\mu.\]
Proof: The sufficiency follows immediately from the expansion formula
\[ e^{itA/\mu}Be^{-itA/\mu} = B + \frac{it}{\mu}[A, B] + \frac{1}{2!} \left( \frac{it}{\mu} \right)^2 [A, [A, B]] + \cdots. \]

We show the necessity. Since the left-hand side of (24) is an equi-spectrum transformation, the right-hand side is also so only when \( B \) has continuous spectrum, so that the Hilbert space \( \mathcal{H} \) on which \( B \) acts must be infinite dimensional. Denoting the eigen-equation for \( B \) by
\[ B|b\rangle = b|b\rangle, \quad (|b\rangle \in \mathcal{H}, \ b \in \mathbb{R}). \quad (25) \]

Then, from (24), we have
\[ e^{if(t)A}Be^{-if(t)A}|b\rangle = g(t)(b - t)|b\rangle, \]
or
\[ Be^{-if(t)A}|b\rangle = g(t)(b - t)e^{-if(t)A}|b\rangle, \]
which indicates that \( e^{-if(t)A}|b\rangle \) is an eigenvector of \( B \) with eigenvalue \( g(t)(b - t) \). On the other hand, operating \( e^{if(t)A} \) to (25) from the left, and using (24), we have
\[ e^{if(t)A}Be^{-if(t)A}e^{if(t)A}|b\rangle = b e^{if(t)A}|b\rangle = g(t)(B - t)e^{if(t)A}|b\rangle, \]
which leads to
\[ Be^{if(t)A}|b\rangle = \left( \frac{b}{g(t)} + t \right)e^{if(t)A}|b\rangle. \]

This indicates that \( e^{if(t)A}|b\rangle \) is an eigenvector of \( B \) with eigenvalue \( (b/g(t) + t) \). Now, \( e^{-if(t)A}|b\rangle \) and \( e^{if(t)A}|b\rangle \) are one-parameter family of eigenvectors of \( B \) which start from a common eigenvector \( |b\rangle \) and, since \( f(t) \) is assumed odd, these eigenvectors must be related by
\[ e^{-if(t)A}|b\rangle = e^{if(-t)A}|b\rangle. \]

Therefore, the corresponding eigenvalues must be identical:
\[ g(t)(b - t) = \left( \frac{b}{g(-t)} - t \right), \]
or
\[ b \left( g(t) - \frac{1}{g(-t)} \right) - t\{g(t) - 1\} = 0. \]

Since this relation must hold for any \( b \) and \( t \), we have \( g(t) = 1 \). In this case, (24) is reduced to
\[ e^{if(t)A}Be^{-if(t)A} = B - t. \]
Expanding the left-hand side as

\[ B + i f(t)[A, B] + \frac{(i f(t))^2}{2!} [A, [A, B]] + \cdots = B - t, \]

and differentiating by \( t \) and setting \( t = 0 \), we have \([A, B] = i/f'(0)\), since \( f(0) = 0 \). Substituting this commutation relation into the above expansion, we see \( f(t) = f'(0)t \). Setting \( \mu = 1/f'(0) \), we have the conditions in the lemma.

These lemmas immediately lead us to the following main theorem.

**Theorem 10:** The one-parameter unitary model (22) becomes e-geodesic iff there exist non-zero reals \( \mu \) and \( J \) such that

\[ f(t) = \frac{t}{\mu}, \quad [A, D_0A] = i\mu^2 J \]

holds. In this case, \( T = D_0A/\mu J \) is the efficient estimator for the parameter \( t \), and \( J \) is the SLD–Fisher information of the model.

**Proof:** From Lemma 9, the conditions derived in Lemma 8 are rewritten as

\[ f(t) = \frac{t}{\mu}, \quad \frac{J(t)}{J(0)} = 1, \quad [A, T] = i\mu, \quad T = \frac{1}{\mu J(0)} D_0A. \]

Setting \( J = J(0) \), we have the theorem.

This theorem shows that, only when the generator is a canonical observable, the one parameter unitary model becomes an e-geodesic, i.e., it has an efficient estimator. Note that this fact is strongly indebted to the assumption that the model is strictly positive. For instance, this condition can be considerably loosened for pure state models, see the next section.

Let us further determine \( \rho_0 \) which satisfy the relation

\[ D_0A = \mu JB, \quad [A, B] = i\mu. \]

Taking the non-commutative Fourier transformation (see Appendix A) of the identity

\[ \rho_0A - A\rho_0 = \frac{i}{2} \mu J[\rho_0B + B\rho_0], \]

we have

\[ -x f_{x,k}(\rho_0) = \mu^2 J \frac{\partial}{\partial x} f_{x,k}(\rho_0). \]
Integrating this under the condition $F_{0,0}\{\rho_0\} = 1$, we have

$$F_{x,k}\{\rho_0\} = \exp(-\frac{x^2}{2\mu^2 J})F(k),$$

where $F(k)$ is an arbitrary function which satisfies $F(0) = 1$ and some regularity conditions. Then $\rho_0$ is written in the Weyl representation as

$$\rho_0 = \int \exp(-\frac{x^2}{2\mu^2 J})F(k) \exp\left[-\frac{i}{\mu}(kA + xB)\right] \frac{dxdk}{2\pi\mu}. \quad (26)$$

For instance, it becomes a quantum Gaussian state if we take $F(k)$ to be Gaussian, see Appendix A.

**Corollary 11:** Quantum Gaussian model is e-autoparallel.

**Proof:** For a Gaussian model, SLDs are

$$L_q = (Q - q)/\sigma_q^2, \quad L_p = (P - p)/\sigma_p^2,$$

the commutation operator

$$\mathcal{D}Q = h(P - p)/\sigma_p^2, \quad \mathcal{D}P = -h(Q - q)/\sigma_q^2,$$

and the SLD Fisher metric

$$J = \begin{bmatrix} 1/\sigma_p^2 & 0 \\ 0 & 1/\sigma_q^2 \end{bmatrix},$$

as shown in Appendix A. The restricted one-dimensional sub-model in the direction $\mathbf{v} = (v_q, v_p)$ is then determined by the SLD

$$S = v_jJ^{ji}L_i = v_q(Q - q) + v_p(P - p) = \frac{1}{\hbar} \mathcal{D}R,$$

where $R = -v_q\sigma_q^2 P + v_p\sigma_p^2 Q$. Note $R$ is independent of $q$ and $p$. The sub-model is therefore written also in the form

$$\rho_t = e^{iRt/\hbar} \rho_0 e^{-iRt/\hbar}.$$

Observing

$$[R, \mathcal{D}_0 R] = i\hbar^2 \left(v_q^2\sigma_q^2 + v_p^2\sigma_p^2\right)$$

we can conclude, with the help of Theorem 10, that the sub-model is an e-geodesic. Since $\mathbf{v}$ is arbitrary, a Gaussian model is proved to be totally e-geodesic. Then it is also e-autoparallel since the model is torsion free. $\blacksquare$
Here we give another example in the spin $2 \times 2$ representation. An e-geodesic (quantum exponential family) which has an efficient estimator $\sigma_z$ is written as [22]
\[
\rho_t = e^{\frac{1}{2}t[\sigma_z - \gamma(t)]} \rho_0 e^{\frac{1}{2}t[\sigma_z - \gamma(t)]}, \quad \gamma(t) = \log[\text{Tr} \rho_0 e^{t\sigma_z}]. \quad (27)
\]
If we set
\[
\rho_0 = \frac{1}{2} \begin{bmatrix} 1 & x_0 \\ x_0 & 1 \end{bmatrix}, \quad -1 \leq x_0 \leq 1
\]
without loss of generality, then (27) becomes
\[
x(t) = \frac{x_0}{\cosh t}, \quad y(t) = 0, \quad z(t) = \tanh t
\]
in the Stokes representation, i.e., $\rho_t = \frac{1}{2}(I + x(t)\sigma_x + y(t)\sigma_y + z(t)\sigma_z)$. It indicates that the e-geodesic is an ellipse of the form
\[
\left\{ \frac{x(t)}{x_0} \right\}^2 + \{z(t)\}^2 = 1,
\]
which connects the north and south poles. Therefore, it cannot be a unitary model unless $x_0 = \pm 1$, i.e., unless $\rho_0$ is a pure state, since the equi-eigenvalue surfaces are spherical shells centered at the origin.

2.4. Conclusions of Section 2
An information geometrical aspects of quantum statistical models were studied. We first introduced a natural dualistic structure on the quantum state space based on the symmetric logarithmic derivatives. We next investigated the autoparallelity and clarified the difference between the classical and the quantum estimation theory. The importance of canonical observables in quantum estimation theory was also stressed.

3. Pure State Estimation Theory
In order to avoid singularities, the conventional quantum estimation theory has been often restricted to models that are composed of strictly positive density operators. Quantum information theorists have also kept away from degenerate states, such as pure states, for mathematical convenience [28]. Indeed, the von Neumann entropy cannot distinguish pure states, and the relative entropies diverge.

In this section, however, we try to construct an estimation theory for pure state models, and clarify the differences between the pure state case
and the strictly positive state case [7, 8, 9]. In Section 3.1, we prove some crucial lemmas which will provide fundamentals of the pure state estimation theory. In Section 3.2, we study the quantum counterpart of the logarithmic derivative and the Fisher information which played important roles in the classical estimation theory. The quantum statistical significance of the Fubini–Study metric is also clarified. In Section 3.3, we provide one parameter pure state estimation theory based on the symmetric logarithmic derivative. This is rather analogous to the conventional quantum estimation theory, but reveals the essential difference between the pure state models and the strictly positive models. In Section 3.4, we study a multi-parameter quantum estimation theory based on the right logarithmic derivative. The estimation theoretical significance of the coherent models is also clarified. In order to demonstrate the results, some examples are also presented.

3.1. Preliminaries

Given a possibly degenerate density operator \( \rho \) on \( \mathcal{H} \), we define sesquilinear forms on \( \mathcal{L} \):

\[
(A, B)_\rho = \text{Tr} \rho BA^*,
\]

\[
\langle A, B \rangle_\rho = \frac{1}{2} \text{Tr} \rho (BA^* + A^*B),
\]

where \( A, B \in \mathcal{L} \). These are apparently the same as the inner products (10) and (11), but are now pre-inner products on \( \mathcal{L} \) that possess all properties of inner products except that \((K, K)_\rho\) may be equal to zero for a nonzero \( K \in \mathcal{L} \). Note that the Schwarz inequality also holds for pre-inner products. The forms \((\cdot, \cdot)_\rho\) become inner products if and only if \( \rho > 0 \). If rank \( \rho = 1 \) or equivalently \( \rho^2 = \rho \), \( \rho \) is called pure. The following lemmas are fundamental.

**Lemma 12:** Suppose \( \rho \) is pure. Then the following three conditions for linear operators \( K \in \mathcal{L} \) are equivalent.

(i) \((K, K)_\rho = 0\),

(ii) \( \rho K = 0 \),

(iii) \( \text{Tr} \rho K = 0 \) and \( \rho K + K^*\rho = 0 \).

**Proof:** Let us express as \( \rho = |\psi\rangle\langle \psi| \) where \( |\psi\rangle \) is a normalized vector in \( \mathcal{H} \). Then the following equivalent sequence

\[
(K, K)_\rho = 0 \iff \langle \psi | KK^* | \psi \rangle = 0 \iff \langle \psi | K = 0 \iff |\psi\rangle \langle \psi | K = 0,
\]
yield (i)⇔(ii). Further, (ii)⇒(iii) is trivial. (iii)⇒(ii) is shown as follows. Operating ⟨ψ| from the left to the assumption

\[ |ψ⟩⟨ψ| K + K^*|ψ⟩⟨ψ| = 0, \]

and invoking another assumption Tr ρK = 0 ⇔ ⟨ψ|K^*|ψ⟩ = 0, we have

\[ 0 = ⟨ψ|ψ⟩⟨ψ| K + ⟨ψ|K^*|ψ⟩⟨ψ| = ⟨ψ|K. \]

Therefore |ψ⟩⟨ψ|K = 0.

\textbf{Lemma 13:} Suppose ρ is pure. Then the following three conditions for linear operators K ∈ L are equivalent.

(i) \( ⟨K, K⟩_ρ = 0, \)
(ii) \( ρK = ρK^* = 0, \)
(iii) \( \text{Tr } ρK = 0, \ ρK + K^*ρ = 0, \text{ and } ρK^* + Kρ = 0. \)

\textit{Proof:} (i)⇔(ii) is shown as follows:

\[ ⟨K, K⟩_ρ = 0 ⇔ ⟨ψ|KK^*|ψ⟩ + ⟨ψ|K^*K|ψ⟩ = 0 \]
\[ ⇔ ⟨ψ|K = 0, ⟨ψ|K^* = 0 \]
\[ ⇔ ρK = ρK^* = 0. \]

(ii)⇔(iii) is a straightforward consequence of Lemma 12.

\textbf{Lemma 14:} Suppose ρ is pure. Then the following three conditions for self-adjoint operators K ∈ L_{sa} are equivalent.

(i) \( ⟨K, K⟩_ρ = 0, \)
(ii) \( ρK = 0, \)
(iii) \( ρK + Kρ = 0. \)


These lemmas are effectively employed in the pure state estimation theory. Denote by \( K(ρ) \) the set of linear operators K ∈ L satisfying \( ⟨K, K⟩_ρ = 0, \) which are called the kernel of the pre-inner product \( ⟨·, ·⟩_ρ. \) Also denote by \( K_{sa}(ρ) \) the set of self-adjoint operators K ∈ L_{sa} satisfying \( ⟨K, K⟩_ρ = 0, \) which are called the kernel of the pre-inner product \( ⟨·, ·⟩_ρ. \)

We next consider Holevo’s commutation operator \( \mathcal{D}, \) which is defined in the same way as (23).

\[ i(Xρ − ρX) = \frac{1}{2} ((\mathcal{D}X)ρ + ρ(\mathcal{D}X)), \quad X, \mathcal{D}X ∈ L_{sa}. \]
If \( \rho \) is degenerate, \( \mathcal{D} X \) is not determined uniquely. In particular, the following lemma holds.

**Lemma 15:** Suppose \( \rho \) is pure. Then \( \mathcal{D} \) is regarded as a super-operator on the quotient space \( L_{sa} / K_{sa}(\rho) \), and is defined by

\[
(\mathcal{D} X) \rho = 2i(X - \text{Tr} \rho X)\rho, \quad (X \in L_{sa} / K_{sa}(\rho)).
\]

**Proof:** Let us denote two distinct images of \( A \in L_{sa} \) by \((\mathcal{D} A)\) and \((\mathcal{D} A)'\), then \( K = (\mathcal{D} A) - (\mathcal{D} A)' \in L_{sa} \) satisfies \( K \rho + \rho K = 0 \) and, from Lemma 13, \( K \in K_{sa}(\rho) \). Further, observing \( \langle \mathcal{D} K, \mathcal{D} K \rangle_\rho = \langle -\mathcal{D}^2 K, K \rangle_\rho \), \( K \in K_{sa}(\rho) \) implies \( \mathcal{D} K \in K_{sa}(\rho) \). Therefore, \( \mathcal{D} \) is regarded as a super-operator on \( L_{sa} / K_{sa}(\rho) \). Further, re-expressing (30) as

\[
\rho \left[ \frac{\mathcal{D} X}{2} + i(X - \text{Tr} \rho X) \right] + \left[ \frac{\mathcal{D} X}{2} + i(X - \text{Tr} \rho X) \right]^* \rho = 0,
\]

and using Lemma 12 together with the identity \( \text{Tr} \rho (\mathcal{D} X) = 0 \), we have an equivalent equation (31).

### 3.2. Quantum Fisher metric

Suppose we are given an \( n \)-parameter pure state model:

\[
S = \{ \rho_\theta : \rho_\theta^* = \rho_\theta, \text{ Tr } \rho_\theta = 1, \rho_\theta^2 = \rho_\theta, \theta \in \Theta \subset \mathbb{R}^n \}. \quad (32)
\]

As in the strictly positive case, we define a family of quantum analogues of the logarithmic derivative by

\[
\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2}[\rho_\theta L_{\theta,j} + L_{\theta,j}^* \rho_\theta], \quad \text{Tr } \rho_\theta L_{\theta,j} = 0. \quad (33)
\]

For instance,

\[
\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2}[\rho_\theta L_{\theta,j}^S + L_{\theta,j}^S \rho_\theta], \quad L_{\theta,j}^S = L_{\theta,j}^{S*} \quad (34)
\]

defines the symmetric logarithmic derivative (SLD). Furthermore, since every pure state model is written in the form \( \rho_\theta = U_\theta \rho_0 U_\theta^* \), where \( U_\theta \) is unitary, we have another useful logarithmic derivative

\[
\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2}[\rho_\theta L_{\theta,j}^A - L_{\theta,j}^A \rho_\theta], \quad \text{Tr } \rho_\theta L_{\theta,j}^A = 0, \quad L_{\theta,j}^A = -L_{\theta,j}^{A*} \quad (35)
\]

which may be called the anti-symmetric logarithmic derivative (ALD). Indeed, the ALD is closely related to the local generator \( A_{\theta,j} = \)
−i(∂Uθ/∂θj)Uθ∗ of the unitary Uθ such as Lθ,j = −2iAθ,j. Thus, (33) defines a certain family of logarithmic derivatives [22]. Denote by T(ρθ) the linear span (over \( \mathbb{R} \)) of logarithmic derivatives which satisfy (33).

**Lemma 16:** Suppose \( \rho_\theta \) is pure and an arbitrary linear operator \( A \in \mathcal{L} \) is given. Then all the quantities \( (A, L_{\theta,j})_{\rho_\theta} \) are identical for every logarithmic derivative \( L_{\theta,j} \in T(\rho_\theta) \) which corresponds to the same direction \( \theta^j \).

**Proof:** Take any logarithmic derivatives \( L_{\theta,j} \) and \( L'_{\theta,j} \) which correspond to the same \( \theta^j \), and denote \( K = L_{\theta,j} - L'_{\theta,j} \). Then, from (33), \( K \) satisfies the condition (iii) of Lemma 12. Therefore \( (K, K)_{\rho_\theta} = 0 \) holds. This and the Schwarz inequality

\[
|\langle A, K \rangle_{\rho_\theta}|^2 \leq (A, A)_{\rho_\theta} (K, K)_{\rho_\theta}
\]

lead to \( (A, K)_{\rho_\theta} = 0 \) for all \( A \in \mathcal{L} \).

From Lemma 16, we can define uniquely the complex Fisher information matrix \( J_\theta \) for the family of logarithmic derivatives (33) whose \((j,k)\) entry is \((L_{\theta,j}, L_{\theta,k})_{\rho_\theta}\).

The SLD is also not uniquely determined for pure state models. Denote by \( T^S(\rho_\theta) \) the linear span of SLDs which satisfy (34).

**Lemma 17:** Suppose \( \rho_\theta \) is pure and an arbitrary self-adjoint operator \( A \in \mathcal{L}_{sa} \) is given. Then all the quantities \( \langle A, L^S_{\theta,j} \rangle_{\rho_\theta} \) are identical for every SLD \( L^S_{\theta,j} \in T^S(\rho_\theta) \) which corresponds to the same direction \( \theta^j \).

**Proof:** By using Lemma 14, it is proved in the same way as Lemma 16.

From Lemma 17, we can define uniquely the real Fisher information matrix \( J^S_\theta \) for the family of SLDs (34) whose \((j,k)\) entry is \( \langle L^S_{\theta,j}, L^S_{\theta,k} \rangle_{\rho_\theta} \), which is called the SLD–Fisher information matrix. The above results are summarized by the following theorem.

**Theorem 18:** Suppose \( \rho_\theta \) is pure. Then the complex Fisher information matrix \( J_\theta = [(L_{\theta,j}, L_{\theta,k})_{\rho_\theta}] \) and the SLD–Fisher information matrix \( J^S_\theta = [\langle L^S_{\theta,j}, L^S_{\theta,k} \rangle_{\rho_\theta}] \) are uniquely determined on the quotient spaces \( T(\rho_\theta)/\mathcal{K}(\rho_\theta) \) and \( T^S(\rho_\theta)/\mathcal{K}_{sa}(\rho_\theta) \), respectively. They are related by \( J^S_\theta = \text{Re} J_\theta \). The \((j,k)\) entry of \( J^S_\theta \) becomes

\[
(J^S_\theta)_{jk} = 2\text{Tr}(\partial_j \rho_\theta)(\partial_k \rho_\theta),
\]

where \( \partial_j = \partial/\partial \theta^j \). This metric is identical, up to a constant factor, to the Fubini–Study metric.
Proof: We only need to prove (36). Differentiating $\rho = \rho^2$,

$$\partial_j \rho = (\partial_j \rho) \rho + \rho (\partial_j \rho).$$  \hspace{1cm} (37)

This relation indicates that $2 \partial_j \rho$ is a representative of the SLD. Then

$$(J^S_{\theta})_{jk} = \langle 2 \partial_j \rho \theta, 2 \partial_k \rho \theta \rangle_{\rho_0} = 2 \text{Tr} \rho_0 [ (\partial_j \rho \theta)(\partial_k \rho \theta) + (\partial_k \rho \theta)(\partial_j \rho \theta) ].$$  \hspace{1cm} (38)

Further, multiplying $\rho_0$ with (37), we have

$$\rho_0 (\partial_j \rho \theta) \rho_0 = 0.$$  \hspace{1cm} (39)

Therefore, by using (37) and (39),

$$\partial_j \rho \theta)(\partial_k \rho \theta) = [(\partial_j \rho \theta) \rho_0 + \rho_0 (\partial_j \rho \theta)][(\partial_k \rho \theta) \rho_0 + \rho_0 (\partial_k \rho \theta)]$$

$$= (\partial_j \rho \theta) \rho_0 (\partial_k \rho \theta) + \rho_0 (\partial_j \rho \theta)(\partial_k \rho \theta).$$

This, along with (38), leads to relation (36). Denoting $\rho_0 = |\psi \rangle \langle \psi |$

$$\text{Tr} (\partial_j \rho \theta)(\partial_k \rho \theta) = 2 [ \text{Re} \langle \partial_j \psi | \partial_k \psi \rangle + \langle \psi | \partial_j \psi \rangle \langle \psi | \partial_k \psi \rangle],$$

which is identical to the Fubini–Study metric [1, 20].

The Fubini–Study metric is known as a gauge invariant metric on a projective Hilbert space [29]. Theorem 18 gives another meaning of the Fubini–Study metric, i.e., the statistical distance. Wootters [31] also investigated from a statistical viewpoint the distance between two rays, and obtained $d(\psi, \varphi) = \cos^{-1} |\langle \psi | \varphi \rangle|$. This is identical, up to a constant factor, to the geodesic distance as measured by the Fubini–Study metric [3]. Theorem 18, together with Theorem 19 below, reveals a deeper connection between them.

3.3. One parameter pure state estimation

In this section, we give a parameter estimation theory of pure state models based on the SLD. As is the strictly positive model case, the following theorem holds.

Theorem 19: Given a pure state model $\rho_0$, the following inequality holds for any locally unbiased measurement $M$:

$$V_0[M] \geq (J^S_{\theta})^{-1}. \hspace{1cm} (40)$$

Proof: It is proved in almost the same way as the strictly positive case (p. 274 [17]), except that $\langle \cdot, \cdot \rangle_{\rho_0}$ is a pre-inner product now. ■
When the model is one dimensional, the measurement $M$ is identified with a certain self-adjoint operator $T$, and the inequalities in the theorem become scalar, i.e.,

$$V_{\theta}[T] \geq \frac{1}{J_{\theta}^{S}}.$$  

(41)

In this case, the lower bound $1/\text{Tr } \rho_\theta (L_\theta^S)^2$ can be attained by the unbiased estimators

$$T = \theta I + \frac{2}{J_{\theta}^{S}} \frac{d\rho_\theta}{d\theta} + K_\theta, \quad \forall K_\theta \in \mathcal{K}_{sa}(\rho_\theta),$$  

(42)

where $I$ is the identity. Since $d\rho_\theta/d\theta$ and $K_\theta$ do not commute in general, the measurement which attains the lower bound (41) is not determined uniquely. This fact provides significant features in the pure state estimation theory which are demonstrated in the following examples.

3.3.1. Time–energy uncertainty relation

Let us consider a model of the form

$$\rho_\theta = e^{i\theta \mathcal{H}/\hbar} \rho_0 e^{-i\theta \mathcal{H}/\hbar}. $$

Here, $\mathcal{H}$ is the time independent Hamiltonian of the system, $\hbar$ the Planck constant, and $\theta$ the time parameter.

According to the one parameter estimation theory for strictly positive models [22],

$$V[T] \geq \frac{1}{J_{\theta}^{S}} \geq \frac{1}{J(L_\theta)} $$

(43)

holds, where $L_\theta$ is any logarithmic derivative which satisfies (33) and $J(L_\theta) = (L_\theta, L_\theta)_{\rho_0}$. The first equality is attained when and only when

$$T = \theta I + L_\theta^S / J_{\theta}^{S}, \quad \text{and the second equality holds iff } L_\theta = L_\theta^S. $$

Now, $L_\theta^A = -2i\mathcal{H}/\hbar$ is an ALD for the model and the corresponding Cramér-Rao inequality becomes

$$V_{\theta}[T] \geq \frac{\hbar^2}{4V_\theta[\mathcal{H}]} $$

(44)

where $T$ is an arbitrary unbiased estimator $T$ for the time parameter $\theta$. This inequality is nothing but a time-energy uncertainty relation. If $\rho_0 > 0$, then this lower bound cannot be attained for any $T$ since $L_\theta^A$ is not an SLD, whereas Theorem 2 asserts that, if $\rho_0$ is pure, the equality in (44) is
locally attainable. This is a significant difference between the strictly positive models and the pure state models. Since the ALD $L^A_\theta = -2i\mathcal{H}/\hbar$ and the SLD–Fisher information for the pure state models $J^S_\theta = 2\text{Tr}(d\rho_\theta/d\theta)^2$ are both obtainable directly from the Liouville–von Neumann equation, this result is not specific to the case where the Hamiltonian is time independent, but is quite general.

### 3.3.2. Efficient estimator

An unbiased estimator $T$ is called efficient if the equality in (41) holds for all $\theta \in \Theta$. Nagaoka [22] has proved that a one parameter model $\rho_\theta$ has an efficient estimator when and only when the model takes the form

$$\rho_\theta = e^{\frac{1}{2}[\beta(\theta)T-\gamma(\theta)]} \rho_0 e^{\frac{1}{2}[\beta(\theta)T-\gamma(\theta)]},$$

where $\beta(\theta), \gamma(\theta)$ are real functions.

Let us consider a model of the form

$$\rho_\theta = e^{if(\theta)A} \rho_0 e^{-if(\theta)A},$$

where $f(\theta)$ is a real monotonic odd function and $A \in \mathcal{L}_{sa}$. If $\rho_0 > 0$, then it was shown that there exists an efficient estimator for $\theta$ only when $A$ is a canonical observable, see Theorem 10. On the other hand, if $\rho_0$ is pure, then there may exist an efficient estimator even if $A$ is not canonical, because of the uncertainty $K_\theta \in \mathcal{K}_{sa}(\rho_\theta)$ in (42). For instance, the spin 1/2 model

$$f(\theta) = \frac{1}{2} \left( \frac{\pi}{2} - \cos^{-1} \theta \right), \quad A = \sigma_y, \quad \rho_0 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

has an efficient estimator $\sigma_z$ for the parameter $\theta$. Indeed, this model admits another form

$$\rho_\theta = \sqrt{1-\theta^2} \exp \left[ \frac{1}{4} \log \frac{1+\theta}{1-\theta} \sigma_z \right] \rho_0 \exp \left[ \frac{1}{4} \log \frac{1+\theta}{1-\theta} \sigma_z \right].$$

This is not a paradox since, in the pure state model, the estimator which attains the Cramér–Rao bound (41) is adjustable for every points $\rho_\theta$ up to the uncertainty of the kernel $\mathcal{K}_{sa}(\rho_\theta)$.

### 3.4. Multi-parameter pure state estimation

When the dimension $n \geq 2$, the matrix equality in (40) cannot be attained in general, because of the impossibility of the exact simultaneous measurement of non-commuting observables (in von–Neumann’s sense). We must, therefore, adopt another strategy to minimize $\text{tr} G\nu_\theta[M], \quad (G \geq 0)$. Since
there is no prototype for general theory of quantum multi-parameter estimation even in the strictly positive case, let us restrict ourselves here to seeking the estimation theory based on the RLD. This may sound strange since the RLD defined by (9) does not always exist for degenerate states. However, it is essential to notice that what we need is not the RLD itself but the inverse of the RLD–Fisher information matrix, as is understood by (13).

We start with the following theorem, which is an extension of Holevo’s result originally obtained in the strictly positive case (p. 280 [17]). Hereafter, the subscripts $\theta$ of the SLDs are omitted for simplicity.

**Theorem 20:** Suppose we are given a pure state model $\rho_\theta$. Let $\{\rho_\theta(\varepsilon) : \varepsilon > 0\}$ be a family of strictly positive density operators $\rho_\theta(\varepsilon)$ having a parameter $\varepsilon$ which satisfy

$$\lim_{\varepsilon \to 0} \rho_\theta(\varepsilon) = \rho_\theta,$$

and denote the corresponding RLD by $L^R_\theta(\varepsilon)$. If the SLD–tangentspace $T^S(\rho_\theta)/K_{sa}(\rho_\theta)$ is $D$-invariant, then

$$\lim_{\varepsilon \to 0} \left( J^R(\varepsilon) \right)^{-1} = \left( J^S \right)^{-1} + i \frac{1}{2} \left( J^S \right)^{-1} D \left( J^S \right)^{-1}$$

holds, where $J^R(\varepsilon) = \left[ \langle L^R_j(\varepsilon), L^R_k(\varepsilon) \rangle_{\rho_\theta(\varepsilon)} \right]$ and $D = [i \text{Tr} \rho_\theta[L^S_j, L^S_k]]$.

**Proof:** Observing the identities

$$\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2} (\rho_\theta L^S_j + L^S_j \rho_\theta) = (L^R_j)^* \rho_\theta$$

and

$$(A, B)_{\rho} = \langle A, (I + \frac{i}{2} D)B \rangle_{\rho},$$

we have, for all $X \in \mathcal{L}$,

$$\langle L^S_j, X \rangle_{\rho_\theta} = \langle L^R_j, X \rangle_{\rho_\theta} = \langle (I + \frac{i}{2} D) L^S_j, X \rangle_{\rho_\theta}.$$

Then $L^S_j = (I + \frac{i}{2} D)L^R_j$ and

$$J^R(\varepsilon) = \left[ \langle L^S_j(\varepsilon), (I + \frac{i}{2} D(\varepsilon))^{-1} L^S_k(\varepsilon) \rangle_{\rho_\theta(\varepsilon)} \right].$$

Since $I + \frac{i}{2} D(\varepsilon)$ is symmetric with respect to $\langle \cdot, \cdot \rangle_{\rho_\theta}$, the following lemma immediately leads us to (47), by setting $V = L_{sa}$, $x_j(\varepsilon) = L^S_j(\varepsilon)$, and $\mathcal{A}(\varepsilon) = I + \frac{i}{2} D(\varepsilon)$.

---

Additional author’s note: There is a crucial error in the proof of Lemma 21. For a modified argument as well as an improved version of Theorem 20, see A. Fujiwara and H. Nagaoka, *J. Math. Phys.*, 40, 4227-4239 (1999), which is reprinted in Chap. 19 of this book.
**Lemma 21:** Let $V$ be a $d$-dimensional linear space (possibly $d = \infty$). Suppose we are given a family of $n(< d)$ linearly independent vectors \( \{x_j(\varepsilon)\}_{j=1}^n \) in $V$, a family of inner products \( \langle \cdot, \cdot \rangle_\varepsilon \), and a family of symmetric operators \( \mathfrak{A}(\varepsilon) \) on $V$ with respect to the inner product, having a parameter \( \varepsilon \geq 0 \). Assume \( \mathfrak{A}(\varepsilon) \) is invertible for \( \varepsilon > 0 \) but \( \mathfrak{A}(0) \) is not. Further, the linear span of \( \{x_j(0)\}_{j=1}^n \) is \( \mathfrak{A}(0) \)-invariant in $V$. Denote $n \times n$ matrices

\[
J^R(\varepsilon) = \begin{bmatrix} \langle x_j(\varepsilon), \mathfrak{A}^{-1}(\varepsilon)x_k(\varepsilon) \rangle_\varepsilon \end{bmatrix}, \quad J(\varepsilon) = \langle x_j(\varepsilon), x_k(\varepsilon) \rangle_\varepsilon.
\]

Then

\[
\lim_{\varepsilon \downarrow 0} (J^R(\varepsilon))^{-1} = J^{-1}(0) \left[ \langle x_j(0), \mathfrak{A}(0)x_k(0) \rangle_0 \right] J^{-1}(0). \quad (48)
\]

**Proof:** Let us denote, by $W^\perp(\varepsilon)$, the orthogonal complement of $W(\varepsilon) = \text{span} \{x_j(\varepsilon)\}_{j=1}^n$ with respect to the inner product $\langle \cdot, \cdot \rangle_\varepsilon$ in $V$. Further, let \( \{y_j(\varepsilon)\}_{j=n+1}^d \) be a basis of $W^\perp(\varepsilon)$ and construct a basis $\{z_j(\varepsilon)\}_{j=1}^d$ of $V$ by combining them as

\[
z_j(\varepsilon) = \begin{cases} x_j(\varepsilon), & j = 1, \cdots, n, \\ y_j(\varepsilon), & j = n + 1, \cdots, d. \end{cases}
\]

Consider enlarged $d \times d$ matrices

\[
\mathfrak{J}^R(\varepsilon) = \begin{bmatrix} \langle z_j(\varepsilon), \mathfrak{A}^{-1}(\varepsilon)z_k(\varepsilon) \rangle_\varepsilon \end{bmatrix}, \quad \mathfrak{J}(\varepsilon) = [\langle z_j(\varepsilon), z_k(\varepsilon) \rangle_\varepsilon].
\]

Since $V = W(\varepsilon) \oplus W^\perp(\varepsilon)$ is, of course, \( \mathfrak{A}(\varepsilon) \)-invariant, the inverse of $\mathfrak{J}^R(\varepsilon)$ is explicitly given as

\[
(\mathfrak{J}^R(\varepsilon))^{-1} = \mathfrak{J}^{-1}(\varepsilon) \left[ \langle z_j(\varepsilon), \mathfrak{A}(\varepsilon)z_k(\varepsilon) \rangle_\varepsilon \right] \mathfrak{J}^{-1}(\varepsilon).
\]

This matrix is well-defined even for $\varepsilon = 0$. Decompose \( (\mathfrak{J}^R(\varepsilon))^{-1} \) into blocks:

\[
\begin{bmatrix} P & O \\ O & Q \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P & O \\ O & Q \end{bmatrix} = \begin{bmatrix} PAP & PBQ \\ QCP & QDP \end{bmatrix},
\]

where the three matrices in the left-hand side correspond to

\[
\mathfrak{J}^{-1}(\varepsilon), \quad [\langle z_j(\varepsilon), \mathfrak{A}(\varepsilon)z_k(\varepsilon) \rangle_\varepsilon], \quad \mathfrak{J}^{-1}(\varepsilon),
\]

respectively, $P, A$ are $n \times n$ matrices, and $Q, D$ are $(d-n) \times (d-n)$ matrices. Further, it is easy to see that $B = C = O$ for $\varepsilon = 0$, since $W(0)$ is $\mathfrak{A}(0)$-invariant and $\mathfrak{A}(0)$ is symmetric. Then $\lim_{\varepsilon \downarrow 0} (\mathfrak{J}^R(\varepsilon))^{-1}$ becomes a block diagonal matrix, and the limit of the first $n \times n$ block $PAP$ approaches (48).
Note that $\text{Tr} \, \rho [L^S_j, L^S_k]$ in Theorem 20 also independent of the uncertainty of the SLD because of Lemma 14. Therefore, Theorem 20 asserts that the inverse of the RLD–Fisher information matrix can be obtained directly from the SLD, without using the diverging RLD–Fisher information matrix itself. Then, it may be important to investigate the condition for the SLD–tangent space to be $\mathfrak{D}$-invariant. The following theorem characterizes the structure of $\mathfrak{D}$-invariant SLD–tangent space.

**Theorem 22:** The $\mathfrak{D}$-invariant SLD–tangent space $T^S(\rho) / K_{sa}(\rho)$ has an even dimension and is decomposed into direct sum of two-dimensional $\mathfrak{D}$-invariant subspaces. Moreover, by taking an appropriate basis of $T^S(\rho) / K_{sa}(\rho)$, the operation of $\mathfrak{D}$ can be written in the form

$$
\mathfrak{D} \begin{bmatrix} \tilde{L}_1^S \\ \tilde{L}_2^S \\ \tilde{L}_3^S \\ \tilde{L}_4^S \\ \vdots \\ \tilde{L}_{2m-1}^S \\ \tilde{L}_{2m}^S \end{bmatrix} = \begin{bmatrix} 0 & 2 & & & & & \\ & -2 & 0 & & & & \\ & & 0 & 2 & & & \\ & & & -2 & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 & 2 \\ & & & & & & & -2 & 0 \end{bmatrix} \begin{bmatrix} \tilde{L}_1^S \\ \tilde{L}_2^S \\ \tilde{L}_3^S \\ \tilde{L}_4^S \\ \vdots \\ \tilde{L}_{2m-1}^S \\ \tilde{L}_{2m}^S \end{bmatrix}.
$$

*(49)*

**Proof:** Since $\text{Tr} \, \rho (\mathfrak{D} X) = 0$ holds for all $X \in L_{sa} / K_{sa}$, (31) is rewritten as

$$
[-4(X - \text{Tr} \, \rho X)]\rho = 2i[\mathfrak{D} X - \text{Tr} \, \rho (\mathfrak{D} X)]\rho.
$$

Comparing this equation to (31) with $X$ replaced by $\mathfrak{D} X$, we have

$$
\mathfrak{D}^2 X = -4(X - \text{Tr} \, \rho X).
$$

In particular, $\mathfrak{D}^2 X = -4X$ holds for every $X$ which satisfies $\text{Tr} \, \rho X = 0$. We may, therefore, write this relation as $\mathfrak{D}^2 = -4$ on the SLD–tangent space $T^S(\rho) / K_{sa}(\rho)$ for short.

Take an arbitrary element $e_1$ of the SLD–tangent space $T^S(\rho) / K_{sa}(\rho)$, and let $e_2 = \mathfrak{D} e_1$. Then from the assumption, $e_2 \in T^S(\rho) / K_{sa}(\rho)$, and $\mathfrak{D} e_2 = -4e_1$ since $\mathfrak{D}^2 = -4$. Therefore, $S_1(\rho) = \text{span} \{ e_1, e_2 \}$ is a $\mathfrak{D}$-invariant subspace of $T^S(\rho) / K_{sa}(\rho)$ and

$$
T^S(\rho) / K_{sa}(\rho) = S_1(\rho) \oplus S_1(\rho)^\perp,
$$

where $S_1(\rho)^\perp$ is the orthogonal complement of $S_1(\rho)$ with respect to $\langle \cdot , \cdot \rangle_{\rho}$. Repeating the same procedure to $S_1(\rho)^\perp$, we have

$$
T^S(\rho) / K_{sa}(\rho) = S_1(\rho) \oplus S_2(\rho) \oplus \cdots \oplus S_m(\rho).
$$
In particular, \( \dim[T^S(\rho_\theta)/K_{sa}(\rho_\theta)] = 2m \).

We next investigate the structure of two-dimensional \( \mathcal{D} \)-invariant subspace \( S_1(\rho_\theta) = \text{span}\{e_1, e_2\} \). Expressing the operation of \( \mathcal{D} \) in a matrix form

\[
\mathcal{D} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} x & y \\ z & w \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}, \quad (x, y, z, w \in \mathbb{R})
\]

and using \( \mathcal{D}^2 = -4 \), we have

\[
x^2 + yz = -4, \quad y(x + w) = 0, \quad z(x + w) = 0, \quad w^2 + yz = -4.
\]

These equations do not contradict the identities \( \langle e_1, \mathcal{D}e_1 \rangle_\rho = \langle e_2, \mathcal{D}e_2 \rangle_\rho = 0 \) iff \( x + w = 0 \), \( y \neq 0 \), \( z \neq 0 \). In this case

\[
\mathcal{D} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} x & y \\ -(x^2 + 4)/y & -x \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}.
\]

Furthermore, the transformation of the basis

\[
\begin{bmatrix} \tilde{L}_1^S \\ \tilde{L}_2^S \end{bmatrix} = \begin{bmatrix} 2/y & 0 \\ x/y & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}
\]

yields

\[
\mathcal{D} \begin{bmatrix} \tilde{L}_1^S \\ \tilde{L}_2^S \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} \tilde{L}_1^S \\ \tilde{L}_2^S \end{bmatrix}.
\]

Repeating the same procedure to other invariant subspaces, we have the theorem.

**Definition 23:** The basis \( \{\tilde{L}_j^S\}_{j=1}^{2m} \) of the SLD–tangent space \( T^S(\rho_\theta)/K_{sa}(\rho_\theta) \) which is subject to the transformation law (49) is called \( \rho_\theta \)-symplectic.

From Theorem 22, it is sufficient to consider a two-dimensional \( \mathcal{D} \)-invariant SLD tangent space. The following theorem gives the condition for the model to have a two-dimensional \( \mathcal{D} \)-invariant SLD–tangent space at \( \rho_\theta \).

**Theorem 24:** For the pure state model \( \{\rho_\theta = |\theta\rangle\langle\theta|\} \), the following two conditions are equivalent.

1. \( \{\tilde{L}_j^S\}_{j=1,2} \) is a \( |\theta\rangle\langle\theta| \)-symplectic basis.
2. \( (\tilde{L}_1^S + i\tilde{L}_2^S)|\theta\rangle = 0 \).

The linear span of such basis \( \text{span}\{\tilde{L}_1^S, \tilde{L}_2^S\} \) is \( \mathcal{D} \)-invariant.
Proof: We first assume (i). Letting $X = \tilde{L}^S_1$ in (31), we have
$$ (\tilde{L}^S_1 + \frac{i}{2} \mathfrak{D} \tilde{L}^S_1) \rho_\theta = 0. $$
Then, $\mathfrak{D} \tilde{L}^S_1 = 2\tilde{L}^S_2$ yields (ii).

Next we assume (ii). Since $(\tilde{L}^S_1 + i\tilde{L}^S_2) \rho_\theta = 0,$
$$ 2\tilde{L}^S_2 \rho_\theta = 2i\tilde{L}^S_1 \rho_\theta,  
-2\tilde{L}^S_1 \rho_\theta = 2i\tilde{L}^S_2 \rho_\theta. $$
Comparing these equations to (31), we have (i).

Since the condition (ii) in Theorem 24 is similar to the definition of the coherent states in quantum theory [18], we shall make the following definition.

**Definition 25:** The pure state model $\{\rho_\theta = |\theta\rangle \langle \theta|\}$ which satisfy the condition in Theorem 24 is called coherent.

Thus the $\mathfrak{D}$-invariency is equivalent to the coherency of the model. The next fact, a straightforward consequence of Theorem 24, characterizes a global structure.

**Corollary 26:** Consider the pure state model of the form $\rho_\theta = U_\theta \rho_0 U_\theta^*$ where $\{U_\theta\}$ forms a projective unitary group. This model is coherent iff $T^S(\rho_0)/K_{sa}(\rho_0)$ is $\mathfrak{D}$-invariant, i.e., the model has a $\rho_0$-symplectic basis. Indeed, if $\{\tilde{L}^S_j\}_{j=1,2}$ is a $\rho_0$-symplectic basis, then $\{U_\theta \tilde{L}^S_j U_\theta^*\}_{j=1,2}$ becomes a $\rho_0$-symplectic basis.

Now we investigate the RLD-bound for coherent models. From Theorem 22, it is sufficient to consider a two-parameter coherent model. In this case, the RLD-bound can be explicitly obtained by substituting (47) into (16). Thus we have

**Proposition 27:** The RLD-bound for a coherent model $\rho_\theta$ is
$$ C^R = C^S + \sqrt{\det G} \frac{\det J^S}{\det J^S} |\text{Tr} \rho_\theta [L^S_1, L^S_2]|. $$
(50)

Since the identity
$$ |\text{Tr} \rho_\theta [L^S_1, L^S_2]| = \text{Tr} \text{Abs} \rho_\theta [L^S_1, L^S_2]. $$
holds for any pure state models, the above CR-bound is similar to the Nagaoka bound, which has been proved most informative when $\rho_\theta$ is represented in $2 \times 2$ matrix [23]. More strongly, we will show in the next subsection (see Theorem 29) that this CR-bound is most informative for any
pure coherent models, by constructing explicitly a generalized measurement that attains this bound. Thus, the coherent model has a nice property from an estimation theoretical viewpoint. Before leaving this subsection, we give two examples of coherent model.

3.4.1. Canonical coherent states

Let us first consider the family of canonical coherent states $\rho_z = |z\rangle \langle z|$ in a one-dimensional harmonic oscillator with frequency $\omega$, where $z = (\omega q + ip)/2\hbar \in \mathbb{C}$, see [11, 12, 19]. This can be regarded as a two-parameter pure state model which has real parameters $q$ and $p$. It is shown that the representative elements of SLDs are

$$L^S_q = \frac{2\omega}{\hbar} (Q - q), \quad L^S_p = \frac{2}{\hbar \omega} (P - p),$$

and

$$\mathcal{D} L^S_q = 2\omega L^S_p, \quad \mathcal{D} L^S_p = - \frac{2}{\omega} L^S_q.$$

Letting

$$\tilde{L}^S_q = \frac{\hbar}{2} L^S_q = \omega (Q - q), \quad \tilde{L}^S_p = \frac{\hbar \omega}{2} L^S_p = P - p,$$

we have

$$\mathcal{D} \tilde{L}^S_q = 2\tilde{L}^S_p, \quad \mathcal{D} \tilde{L}^S_p = -2\tilde{L}^S_q.$$

This indicates that $\{\tilde{L}^S_q, \tilde{L}^S_p\}$ forms a $\rho_z$–symplectic basis. Therefore, from Theorem 24,

$$(\tilde{L}^S_q + i \tilde{L}^S_p)|z\rangle = [\omega (Q - q) + i(P - p)]|z\rangle = 0,$$

which is nothing but the definition of canonical coherent states. Furthermore, from Theorem 20, we obtain

$$(JR)^{-1} = \begin{bmatrix} \sigma_P^2 & \frac{i\hbar}{2} \\ -\frac{i\hbar}{2} & \sigma_Q^2 \end{bmatrix},$$

where $\sigma_P^2 = \hbar \omega/2$, $\sigma_Q^2 = \hbar/2\omega$, and the corresponding RLD–bound

$$g_P V_P[M] + g_Q V_Q[M] \geq g_P \sigma_P^2 + g_Q \sigma_Q^2 + \hbar \sqrt{g_P g_Q}$$

is identical to the pure state limit of the most informative CR bound obtained by Yuen and Lax [32] (p. 281 [17]).
3.4.2. Spin coherent states

Another example is the family of spin coherent states, see Appendix B. Let \((\theta, \varphi)\) be the polar coordinates where the north pole is \(\theta = 0\) and \(x\)-axis corresponds to \(\varphi = 0\). The spin coherent state \(|\theta, \varphi\rangle\) is defined as

\[
|\theta, \varphi\rangle = R[\theta, \varphi]|j\rangle = \exp\left[i\theta(J_x \sin \varphi - J_y \cos \varphi)\right]|j\rangle,
\]

where \(|j\rangle\) is the highest occupied state in the spin \(j\) system. It is shown that the SLDs at the north pole in the direction of \(\varphi = 0\) and \(\varphi = \pi/2\) are, respectively, \(2J_x\), \(2J_y\) and the operation of \(\mathfrak{D}\) becomes \(\mathfrak{D}J_x = 2J_y\), \(\mathfrak{D}J_y = -2J_x\). Therefore, \(\tilde{L}_1^S = J_x\) and \(\tilde{L}_2^S = J_y\) form a \(|j\rangle\langle j|\)-symplectic basis and from Theorem 24,

\[
\left(\tilde{L}_1^S + i\tilde{L}_2^S\right)|j\rangle = J_+|j\rangle = 0,
\]

where \(J_+ = J_x + iJ_y\) is the spin creation operator. This is nothing but the definition of the terminal state \(|j\rangle\). From this fact, we can immediately conclude that the model which comprises the totality of the spin coherent states

\[
\rho_{\theta, \varphi} = |\theta, \varphi\rangle\langle \theta, \varphi| = R[\theta, \varphi]|j\rangle\langle j|R[\theta, \varphi]^{-1}
\]

has \(\mathfrak{D}\)-invariant SLD tangent space at every point on the sphere. Indeed, since \(R[\theta, \varphi]\) form a compact Lie group, Corollary 26 asserts that

\[
\left\{R[\theta, \varphi]\tilde{L}_1^S R[\theta, \varphi]^{-1}, R[\theta, \varphi]\tilde{L}_2^S R[\theta, \varphi]^{-1}\right\}
\]

form a \(|\theta, \varphi\rangle\langle \theta, \varphi|\)-symplectic basis. In particular, a two-parameter spin 1/2 model has \(\mathfrak{D}\)-invariant SLD tangent space and

\[
\left(\mathfrak{J}^R\right)^{-1} = \frac{1}{\sin^2 \theta} \begin{bmatrix} \sin^2 \theta & -i \sin \theta \\ i \sin \theta & 1 \end{bmatrix}.
\]

The corresponding CR bound is

\[
g_\theta V_\theta[M] + g_\varphi V_\varphi[M] \geq g_\theta + \frac{g_\varphi}{\sin^2 \theta} + \frac{2}{\sin \theta} \sqrt{g_\theta g_\varphi}.
\]

This bound is identical to the pure state limit of the most informative CR bound obtained by Nagaoka [23].

3.5. Linear random measurement

Suppose we are given a two-parameter model \(\{\rho_\theta ; \theta = (\theta_1, \theta_2) \in \mathbb{R}^2\}\), the corresponding SLD being \(\{L_{\theta,1}^S, L_{\theta,2}^S\}\). In the following, we often drop the
Geometry of Quantum Estimation Theory

subscript $\theta$ for notational convenience since we only consider local properties of the model. If $L^S_1$ and $L^S_2$ commute, then we can estimate the parameters $(\theta^1, \theta^2)$ in the same way as in the classical theory. Therefore, suppose $[L^S_1, L^S_2] \neq 0$. If one of the two parameters, say $\theta^2$, is fixed, then we obtain a one-parameter sub-model $\{\rho_{\theta^1 \theta^2} ; \theta^2 = \text{const.}\}$. For this sub-model, we have an optimum locally unbiased estimator $T^1 = \theta^1 I + L^S_1 / (J^S)_{11}$ for the parameter $\theta^1$ as was mentioned in the previous section, where $(J^S)_{11}$ is the $(1,1)$ component of the Fisher information matrix $J^S$. Therefore it is natural to ask whether the optimum estimation for the original two parameter model $\rho_{\theta}$ can be realized by a random measurement of two “observables” $L^S_1$ and $L^S_2$. More generally, let us investigate the infimum of $\text{tr} \, GV[M]$ with respect to the random measurements $M$ of linearly independent two observables $A_1, A_2$ in the linear span

$$\mathcal{L}^S = \{a_1 L^S_1 + a_2 L^S_2 ; a_1, a_2 \in \mathbb{R}\},$$

see Figure 1. Note that if we obtain the optimum measurement $M$ for $G = I$, the corresponding locally unbiased operators being $X^j$, the solution for general weight $G$ is $Y^j = \sum_{k=1}^n f^j_k X^k$, where $\sqrt{G} = [f^j_k]$. We then consider only $G = I$ case in the following without loss of generality.

Denote the dual basis by $L^i = J^{ij} L^S_j$ where $J^{ij}$ is the $(i,j)$ entry of inverse of the SLD–Fisher information matrix. A pair of self-adjoint operators $X^1, X^2$ whose expectations vanish is locally unbiased with respect to the parameters $\theta^1, \theta^2$ iff $\langle L^S_i, X^j \rangle = \delta^i_j$ because of (5), where $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{\rho_{\theta}}$. This condition is rewritten as $\langle L^i, X^j \rangle = \langle L^i, L^j \rangle$, which indicates that $L^i$ is the orthogonal projection of $X^j$ onto the SLD–tangent space with respect to $\langle \cdot, \cdot \rangle$, see Figure 2. We therefore restrict ourselves to the case $X^j = L^j, (j = 1, 2)$. Further, let $A_1, A_2 \in \mathcal{L}^S$ be linearly independent observables which are to be measured at random, assuming $\langle A_j, A_j \rangle = 1$.

Fig. 1. Random measurement of two non-commuting observables. It is expected that the optimal measurement might be given by a random measurement of two “observables” $L^S_1$ and $L^S_2$ which gives the best estimates for $\theta^1$ and $\theta^2$, respectively.
Fig. 2. Geometrical interpretation of local unbiasedness condition. Every locally unbiased operators $X^j$ have the common orthogonal projection $L^j$ which is the dual of $L^S_j$.

without loss of generality. Their spectral decompositions are written as

$$A_j = \sum_\xi a_j(\xi)E_j(\xi).$$

Let us construct their random measurement as follows. Select one of $A_1, A_2$ according to the probability $p_1, p_2$, respectively, and make an exact measurement of it (in von Neumann’s sense). The corresponding resolution of identity is defined by

$$M_j(\xi) = p_jE_j(\xi).$$

When we selected an observable $A_j$ and obtained an outcome $a_j(\xi)$, we identify this result to a pair of real quantities $b^1_j(\xi)$ and $b^2_j(\xi)$, which are connected to $L^1, L^2$ by

$$b^k_j(\xi) = \frac{1}{p_j} \langle L^k, A^j \rangle a_j(\xi), \quad (j,k = 1, 2),$$

where $\{A^1, A^2\}$ is the dual basis of $\{A_1, A_2\}$ in $L^S$ with respect to the inner product $\langle \cdot, \cdot \rangle$. Indeed, the following relation holds:

$$L^k = \sum_{j=1}^2 \sum_\xi b^k_j(\xi)M_j(\xi).$$

Then, by comparing the decomposition (54) to (4), we can evaluate the
equi-weighted trace of covariance matrix as

\[
\text{tr} \, V[M] = \sum_{j=1}^{2} \sum_{\xi} \left[ (b_j^{1}(\xi))^2 + (b_j^{2}(\xi))^2 \right] \text{Tr} \rho M_j(\xi)
\]

\[
= \sum_{j} \frac{1}{p_j} \left[ \langle L^1, A^j \rangle^2 + \langle L^2, A^j \rangle^2 \right].
\]  \hfill (55)

Our aim is to find the infimum of (55) with respect to \(\{p_j\}\) and \(\{A_j\}\).

Observing the fact that \(\mu_1/p_1 + \mu_2/p_2\), \(p_1 + p_2 = 1\) takes its minimum \((\sqrt{\mu_1} + \sqrt{\mu_2})^2\) when and only when \(p_j = \sqrt{\mu_j}/(\sqrt{\mu_1} + \sqrt{\mu_2})\), we have

\[
\min_{\{p_j\}} \text{tr} \, V[M] = \left[ \sqrt{\langle L^1, A^1 \rangle^2} + \sqrt{\langle L^2, A^1 \rangle^2} + \sqrt{\langle L^1, A^2 \rangle^2} + \sqrt{\langle L^2, A^2 \rangle^2} \right]^2.
\]  \hfill (56)

The problem is then reduced to the minimization of (56) with respect to \(A_1, A_2\). The normalization conditions \(\langle A_1, A_1 \rangle = \langle A_2, A_2 \rangle = 1\) impose the following constraints on \(A^1, A^2\),

\[
\langle A^1, A^1 \rangle = \langle A^2, A^2 \rangle = \frac{1}{1 - \alpha^2}, \quad \langle A^1, A^2 \rangle = -\frac{\alpha}{1 - \alpha^2},
\]  \hfill (57)

where \(\alpha = \langle A_1, A_2 \rangle\).

Let us define a linear transformation \(\phi : L^S \rightarrow L^S\) by

\[
\phi(W) = \langle L^1, W \rangle L^1 + \langle L^2, W \rangle L^2.
\]  \hfill (58)

Since \(\phi\) is symmetric and positive definite, it has positive eigenvalues \(\lambda_1, \lambda_2\) and unit eigenvectors \(U_1, U_2\), satisfying

\[
\phi(U_j) = \lambda_j U_j, \quad (j = 1, 2).
\]  \hfill (59)

By expanding as \(A^i = a^{ij} U_j\), the problem to be solved is written in the form

\[
\text{minimize}_{A^1, A^2} \left[ \sqrt{\langle A^1, \phi(A^1) \rangle} + \sqrt{\langle A^2, \phi(A^2) \rangle} \right]^2
\]

\[
= \text{minimize}_{A^1, A^2} \left[ \sqrt{\lambda_1 (a^{11})^2 + \lambda_2 (a^{12})^2} + \sqrt{\lambda_1 (a^{21})^2 + \lambda_2 (a^{22})^2} \right]^2.
\]

The constraints (57) become

\[
(a^{11})^2 + (a^{12})^2 = (a^{21})^2 + (a^{22})^2 = \frac{1}{1 - \alpha^2}, \quad (a^{11})(a^{21}) + (a^{12})(a^{22}) = -\frac{\alpha}{1 - \alpha^2}.
\]
Introduce another parameterization by
\[
a_{11} = \frac{1}{\sqrt{1 - \alpha^2}} \cos \theta, \quad a_{12} = \frac{1}{\sqrt{1 - \alpha^2}} \sin \theta,
\]
\[
a_{21} = \frac{1}{\sqrt{1 - \alpha^2}} \cos \varphi, \quad a_{22} = \frac{1}{\sqrt{1 - \alpha^2}} \sin \varphi.
\]
Then the remaining third constraint becomes \(\cos(\theta - \varphi) = -\alpha\), and we have the constraint free minimization problem as
\[
\min_{\theta, \varphi, \alpha} \frac{1}{1 - \alpha^2} \left[ \sqrt{\lambda_1 \cos^2 \theta + \lambda_2 \sin^2 \theta} + \sqrt{\lambda_1 \cos^2 \varphi + \lambda_2 \sin^2 \varphi} \right]^2
\]
\[
= \min_{\theta, \varphi} \frac{1}{\sin^2(\theta - \varphi)} \left[ \sqrt{a + b \cos 2\theta} + \sqrt{a + b \cos 2\varphi} \right]^2,
\]
where \(a = (\lambda_1 + \lambda_2)/2\), \(b = (\lambda_1 - \lambda_2)/2\). By using the inequality (see Appendix C)
\[
\frac{1}{\sin(\theta - \varphi)} \left[ \sqrt{a + b \cos 2\theta} + \sqrt{a + b \cos 2\varphi} \right] \geq \sqrt{a + b} + \sqrt{a - b}, \quad (60)
\]
which holds for \(0 < \theta - \varphi < \pi\), we obtain the desired infimum
\[
\min_{\{A_j\}} \min_{\{p_j\}} \tr V[M] = \left[ \sqrt{\lambda_1} + \sqrt{\lambda_2} \right]^2
\]
\[
= \langle L^1, L^1 \rangle + \langle L^2, L^2 \rangle + 2\sqrt{\langle L^1, L^1 \rangle \langle L^2, L^2 \rangle - \langle L^1, L^2 \rangle^2}.
\]
(61)
The last equality follows from the fact that the trace \(\lambda_1 + \lambda_2\) and the determinant \(\lambda_1 \lambda_2\) of the linear transformation \(\phi\) is independent of the choice of the basis which represents \(\phi\) in a matrix form. Since the equality in (60) holds on a sinusoidal periodic curve, there exist continuous potency of optimum measurements that attain the minimum (61). The lower bound (61) is first appeared in [25, 26], although its meaning is not clearly stated.

Let us give two applications which clarify the significance of the lower bound (61).

3.5.1. Gaussian model

For a quantum Gaussian model (see Appendix A), the SLDs become
\[
L_q^S = \frac{1}{\sigma_q^2} (Q - q), \quad L_p^S = \frac{1}{\sigma_p^2} (P - p),
\]
where \(\sigma_q^2 = \langle Q - q, Q - q \rangle\) and \(\sigma_p^2 = \langle P - p, P - p \rangle\) are variances of \(Q\) and \(P\), respectively. Then,
\[
J^S = \begin{bmatrix} 1/\sigma_q^2 & 0 \\ 0 & 1/\sigma_p^2 \end{bmatrix}, \quad (J^S)^{-1} = \begin{bmatrix} \sigma_q^2 & 0 \\ 0 & \sigma_p^2 \end{bmatrix},
\]
and the duals of $L_q$ and $L_p$ are

$$L^q = Q - q, \quad L^p = P - p,$$

respectively. Therefore, the bound (61) becomes

$$\text{tr} V[M] = \sigma_q^2 + \sigma_p^2 + 2\sigma_q\sigma_p.$$

This is greater than the most informative bound (see Appendix A)

$$\text{tr} V[M] = \sigma_q^2 + \sigma_p^2 + \hbar$$

unless $\sigma_q\sigma_p = \hbar/2$, i.e., unless the system is in a minimum uncertainty state. Furthermore, even in the pure coherent state case, it is an open question whether the covariant measurement

$$M(p, q) = U[p, q]|0\rangle\langle 0|U^*[p, q],$$

which is known to attain the above most informative bound (p. 281 [17]), can be constructed directly from the continuous combination of the above optimal random measurements.

### 3.5.2. Pure coherent models

In this section, we will prove that the RLD–bound (50) gives the most informative CR–bound for any quantum coherent model by showing that it is attainable by random measurements.

**Definition 28:** The parameter $\theta$ of a model $\rho_\theta$ is called $\rho_0$–canonical if the SLD–Fisher information matrix with respect to $\theta$ is in diagonal form at $\rho_0$.

This condition is not restrictive since, by a certain transformation of coordinate system, we can always diagonalize the SLD–Fisher information matrix.

**Theorem 29:** Suppose the parameter $\theta$ of the coherent model $\rho_\theta$ is $\rho_0$–canonical. Then the corresponding RLD–bound is attainable at $\rho_0$ by a certain random measurement.

**Proof:** After some calculation, we have another expression for the RLD–bound (50) as

$$C^R = \langle L^1, L^1 \rangle + \langle L^2, L^2 \rangle + |\text{Tr} \rho_\theta[L^1, L^2]|,$$

where we set $G = I$. On the other hand, from the assumption, there exists non–zero real numbers $c_1, c_2$ and normalized $\rho_0$–symplectic basis $\{\tilde{L}_1^S, \tilde{L}_2^S\}$
such that $L_j^S = c_j \tilde{L}_j^S$, or $L_j^j = \tilde{L}_j^j / c_j$. Substituting this relation to the condition (ii) in Theorem 24, we have
\[(c_1 L_1^1 + ic_2 L_1^2) \rho_0 = 0,\]
which is nothing but the minimum uncertainty condition in the Heisenberg’s uncertainty relation. Then
\[\langle L_1^1, L_1^1 \rangle \langle L_2^2, L_2^2 \rangle = \frac{1}{4} \left| \text{Tr} \rho_0 [L_1^1, L_2^2] \right|^2, \quad \langle L_1^1, L_2^2 \rangle = 0. \quad (63)\]
Comparing (61) (62) and (63), we have the theorem.

This proof is indebted to the special choice of the coordinate system. It is not yet clear whether $C^R$ can be attained in an arbitrary coordinate system.

Here we give an example of a two parameter coherent model which have non-diagonal SLD–Fisher information matrix, called the photon squeezed state model (see Appendix D). Throughout this example, adjoint operators and complex conjugate numbers are denoted by $\dagger$ and $\ast$, respectively, according to the convention in physics. A photon squeezed state $|z\rangle_{\xi}$ is defined as
\[|z\rangle_{\xi} = D(z)S(\xi)|0\rangle, \quad D(z) = \exp(za^{\dagger} - z^{*}a), \quad S(\xi) = \exp[\frac{1}{2}(\xi a^{\dagger}^2 - \xi^{*} a^2)],\]
where $z, \xi$ are complex numbers, $a$ and $a^{\dagger}$ are the annihilation and creation operators of a boson satisfying $[a, a^{\dagger}] = 1, [a, a] = [a^{\dagger}, a^{\dagger}] = 0$, and $|0\rangle$ is the Fock vacuum with respect to $a$. Letting $z = (q + ip)/\sqrt{2}$, we may regard the family of density operators
\[\rho_z = |z\rangle_{\xi\xi}\langle z| = D(z)\rho_0 D^{\ast}(z), \quad \rho_0 = |0\rangle_{\xi\xi}\langle 0|\]
as a quantum parametric model which is parameterized by two real numbers $q$ and $p$. The pre-inner products with respect to the position operator $Q = (a + a^{\dagger})/\sqrt{2}$ and the momentum operator $P = (a - a^{\dagger})/i\sqrt{2}$ become
\[\begin{bmatrix}
(Q - q, Q - q)_{\rho_z} & (Q - q, P - p)_{\rho_z} \\
(P - p, Q - q)_{\rho_z} & (P - p, P - p)_{\rho_z}
\end{bmatrix}
= \frac{1}{2} \begin{bmatrix}
(\cosh 2s + \cos \theta \sinh 2s) & (\sin \theta \sinh 2s + i) \\
(\sin \theta \sinh 2s - i) & (\cosh 2s - \cos \theta \sinh 2s)
\end{bmatrix},\]
\[\text{Additional author’s note: This problem has been resolved affirmatively in A. Fujiwara and H. Nagaoka, J. Math. Phys., 40, 4227-4239 (1999), which is reprinted as Chap. 19 of this book.}\]
where $\xi = se^{i\theta}$. Noting $\text{Tr} \rho_z L_j^A = 0 \ (j = q, p)$, the ALDs are readily obtained as

$$L_q^A = 2i(P - p), \quad L_p^A = -2i(Q - q).$$

Further, observing the relation

$$b|z\rangle_{\xi} = \beta|z\rangle_{\xi},$$

where

$$b = S(\xi)aS^{-1}(\xi) = acosh s - a^\dagger e^{i\theta} \sinh s,$$

$$\beta = z \cosh s - z^* e^{i\theta} \sinh s,$$

we immediately have a normalized $\rho_z$–symplectic basis as

$$\tilde{L}^S_q = \sqrt{2} \left[ (Q - q)(\cosh s - \cos \theta \sinh s) - (P - p) \sin \theta \sinh s \right],$$

$$\tilde{L}^S_p = \sqrt{2} \left[ (P - p)(\cosh s + \cos \theta \sinh s) - (Q - q) \sin \theta \sinh s \right].$$

On the other hand, from (30) (34) and (35), the SLDs are given as $L_j^S = \mathcal{D}(-L_j^A/2i), \ (j = q, p)$. Then, by expanding $\{L_j^A\}$ into linear combinations of $\{\tilde{L}_j^S\}$, and using the relations $\mathcal{D} \tilde{L}_q^S = 2\tilde{L}_p^S$ and $\mathcal{D} \tilde{L}_p^S = -2\tilde{L}_q^S$, we have

$$L_q^S = 2[ (Q - q)(\cosh 2s - \cos \theta \sinh 2s) - (P - p) \sin \theta \sinh 2s],$$

$$L_p^S = 2[ (P - p)(\cosh 2s + \cos \theta \sinh 2s) - (Q - q) \sin \theta \sinh 2s].$$

Then the SLD–Fisher information matrix becomes

$$J^S = \begin{bmatrix}
\langle L_q^S, L_q^S \rangle_{\rho_z} & \langle L_q^S, L_p^S \rangle_{\rho_z} \\
\langle L_p^S, L_q^S \rangle_{\rho_z} & \langle L_p^S, L_p^S \rangle_{\rho_z}
\end{bmatrix}
= 2 \begin{bmatrix}
\cosh 2s - \cos \theta \sinh 2s & -\sin \theta \sinh 2s \\
-\sin \theta \sinh 2s & \cosh 2s + \cos \theta \sinh 2s
\end{bmatrix}. $$

We thus have the dual basis of $L_q^S, L_p^S$ as

$$L^q = Q - q, \quad L^p = P - p.$$  

Furthermore, this metric is, up to a constant factor, identical to the Fubini–Study metric [1], and is also identical to the real part of the complex ALD–Fisher information matrix

$$J^A = \begin{bmatrix}
\langle L_q^A, L_q^A \rangle_{\rho_z} & \langle L_q^A, L_p^A \rangle_{\rho_z} \\
\langle L_p^A, L_q^A \rangle_{\rho_z} & \langle L_p^A, L_p^A \rangle_{\rho_z}
\end{bmatrix}
= 2 \begin{bmatrix}
\cosh 2s - \cos \theta \sinh 2s & -\sin \theta \sinh 2s + i \\
-\sin \theta \sinh 2s - i & \cosh 2s + \cos \theta \sinh 2s
\end{bmatrix}. $$
as is understood by Theorem 18. This result indicates that the coordinate system \((q, p)\) is a non-orthogonal one. Let us change the coordinate system \((q, p)\) into another one \((q', p')\) such that the corresponding SLD \(L^S_{q'}, L^S_{p'}\) are orthogonal at \(\rho_z\) with respect to the pre-inner product \(\langle \cdot, \cdot \rangle_{\rho_z}\). For instance, the transformation of coordinate system

\[
q' = q(\cosh s - \cos \theta \sinh s) - p \sin \theta \sinh s \\
p' = p(\cosh s + \cos \theta \sinh s) - q \sin \theta \sinh s
\]

lead to \(L^S_{q'} = \sqrt{2}L^S_q, L^S_{p'} = \sqrt{2}L^S_p\). Then, according to Theorem 4.1, there exist random measurements that attain the RLD bound with respect to \((q', p')\). In this example, moreover, \(\min M \text{tr } V[M]\) for the original parameter \((q, p)\) in (61) also coincides with \(C^R\) in (62) as

\[
\min_M \text{tr } V[M] = C^R = \cosh 2s + 1.
\]

3.6. Geometry of pure state estimation

In this section, we make a supplementary consideration toward the information geometry of quantum pure states. This attempt has not been completed and most problems are left for future study. \(^4\)

3.6.1. Dualistic structure of pure state space

We have introduced in Chapter 2 a dualistic structure in the space of strictly positive density operators, where the m-tangent vector and the e-tangent vector are uniquely determined and a pair of tele-parallel translations are naturally defined which make the state space dually curvature free. In the pure state case, however, the e-tangent vector (the SLD) is not uniquely determined, i.e., it is determined up to the kernel \(K_{sa}(\rho) = \{ K; \rho K = 0, K = K^* \}\) as shown in the previous chapter. Moreover, to make matters worse, we cannot even introduce tele-parallel translations, since the set of m-tangent and e-tangent vectors cannot be identified with the set of operators satisfying \(\text{Tr } G = 0\) and \(\text{Tr } \rho H = 0\), respectively. This implies that the pure state space \(\mathcal{P}\) inherently has non-vanishing curvatures even in the dualistic geometry. We must therefore define a dualistic structure by means of a pair of infinitesimal parallel translations, i.e., covariant derivatives. Let us first define m- and e-representations of tangent vectors.

• m-tangent vector
\[ T_\rho(\mathcal{P}) \simeq T_\rho^{(m)}(\mathcal{P}) \]
\[ \partial \simeq \partial \rho \]

• e-tangent vector
\[ T_\rho(\mathcal{P}) \simeq T_\rho^{(e)}(\mathcal{P}) \]
\[ \partial \simeq L \]

Here, \( L \) is the SLD defined by
\[ \partial \rho = \frac{1}{2} [\rho L + L \rho], \quad L = L^*, \]
which is determined up to \( \mathcal{K}_{sa}(\rho) \). As shown in the previous chapter, the Riemannian metric is uniquely defined regardless of the uncertainty of \( L \).
\[ g_\rho(\partial_i, \partial_j) = \langle L_i, L_j \rangle_\rho = \frac{1}{2} \text{Tr} \rho (L_i L_j + L_j L_i) = \text{Tr} (\partial_i \rho) L_j. \]

On the other hand, a pair of covariant derivatives\(^\dagger\) are not uniquely determined due to the uncertainty of \( L \):

• m-covariant derivative
\[ \nabla_{\partial_i} \partial_j = \partial_i \partial_j \rho. \]

• e-covariant derivative
\[ \nabla_{\partial_i}^{(e)} \partial_j = \partial_i L_j - \text{Tr} \rho (\partial_i L_j) + \frac{\delta K_j}{d\theta^i}. \]

Here, \( \delta K_j \) is an arbitrary infinitesimal element in \( \mathcal{K}_{sa}(\rho) \). Note \( \delta K_j/d\theta^i \in \mathcal{K}_{sa}(\rho) \) and it does not mean the derivative of \( K_j(\rho) \). The corresponding connection coefficients become
\[ \Gamma^{(m)}_{ij,k} = \langle \nabla_{\partial_i}^{(m)} \partial_j, \partial_k \rangle_\rho = \text{Tr} (\partial_i \partial_j \rho) L_k, \]
\[ \Gamma^{(e)}_{ij,k} = \langle \nabla_{\partial_i}^{(e)} \partial_j, \partial_k \rangle_\rho = \text{Tr} (\partial_i L_j)(\partial_k \rho), \]
where \( \text{Tr} (\delta K_j/d\theta^i)(\partial_k \rho) = 0 \) is used. These formulae immediately lead to the duality of the connections with respect to the Riemannian metric:
\[ \partial_i g(\partial_j, \partial_k) = g(\nabla_{\partial_i}^{(m)} \partial_j, \partial_k) + g(\partial_j, \nabla_{\partial_i}^{(e)} \partial_k). \]
Thus, the dualistic structure is determined by specifying a smooth family of e-tangent vectors \( \{ L_i(\rho) \} \).

\(^\dagger\)Additional author’s note: These definitions need slight modifications. See the reference cited in the footnote of p. 262.
3.6.2. Curvature and torsion

As stated before, the curvatures for pure state space never vanishes. On the other hand, we can expect that the torsion, which was inevitably nonzero in the strictly positive state case, may vanish in the pure state case by a suitable choice of the smooth family of e-tangent vectors. The torsion tensor is given (p. 46 [2]) by

\[ S_{ij,k}^{(e)} = \Gamma_{ij,k}^{(e)} - \Gamma_{ji,k}^{(e)} = \text{Tr} (\partial_i L_j - \partial_j L_i)(\partial_k \rho). \]

Therefore, if we adopt, for instance, \(2\partial \rho\) as the representative element of the SLD, the torsion tensor indeed vanishes. The next theorem gives the general condition for the torsion tensor to vanish.

**Theorem 30:** Suppose we adopt a smooth family of e-tangent vectors \(\{L_i(\rho)\}\). The corresponding torsion tensor vanishes iff

\[ \partial_i K_j(\rho) - \partial_j K_i(\rho) \in K_{sa}(\rho), \]

where \(K_i(\rho)\) denotes the deviation of e-tangent vector \(L_i(\rho)\) from \(2\partial_i \rho\), i.e.

\[ L_i(\rho) = 2\partial_i \rho + K_i(\rho), \quad K_i(\rho) \in K_{sa}(\rho). \]

**Proof:** The sufficiency is obvious. We show the necessity. For notational simplicity, we drop the argument \(\rho\).

\[
S_{ij,k}^{(e)} = \text{Tr} (\partial_i L_j - \partial_j L_i)(\partial_k \rho)
= \text{Tr} (\partial_i K_j - \partial_j K_i)(\partial_k \rho)
= \frac{1}{2} \text{Tr} (\partial_i K_j - \partial_j K_i)(\rho L_k + L_k \rho)
= \frac{1}{2} \text{Tr} \{(\partial_i K_j - \partial_j K_i)\rho + \rho(\partial_i K_j - \partial_j K_i)\} L_k.
\]

Since \(\rho K = 0\), \(\rho(\partial K) = -(\partial \rho)K\) holds. Substituting this and its conjugate into the above equation, we have

\[
S_{ij,k}^{(e)} = \frac{1}{2} \text{Tr} \{-K_j(\partial_i \rho) + K_i(\partial_j \rho) - (\partial_i \rho)K_j + (\partial_j \rho)K_i\} L_k
= \frac{1}{4} \text{Tr} \{-K_j(\rho L_i + L_i \rho) + K_i(\rho L_j + L_j \rho)\}
- (\rho L_i + L_i \rho)K_j + (\rho L_j + L_j \rho)K_i\} L_k
= \frac{1}{4} \text{Tr} \{(K_i L_j - K_j L_i)\rho L_k + L_k \rho(L_j K_i - L_i K_j)\}
= \frac{1}{2} \text{Re} \langle l_k | \phi_{ij} \rangle,
\]
where $\rho = |\psi\rangle\langle\psi|$, and $|l_k\rangle = L_k|\psi\rangle$, $|\phi_{ij}\rangle = (K_iL_j - K_jL_i)|\psi\rangle$. Since $|l_k\rangle \in \mathcal{H}$ is arbitrary except that $\langle\psi|l_k\rangle = 0$, $S_{ij,k}^{(e)} = 0$ implies $|\phi_{ij}\rangle = 0$ for all $i, j$ (Note also $\langle\psi|\phi_{ij}\rangle = 0$). This is equivalent to $(K_iL_j - K_jL_i)\rho = 0$, or $K_i(\partial_j\rho) = K_j(\partial_i\rho)$, which implies $(\partial_iK_j)\rho = (\partial_jK_i)\rho$ due to $\rho K = 0$. Therefore $\partial_iK_j - \partial_jK_i \in K_{sa}$.

Theorem 30 asserts that, by a suitable choice of the element of the kernel, the torsion of the dualistic structure for pure states may vanish. For instance, the dualistic structure reduces to the Riemannian structure if we adopt $L(\rho) = 2\partial\rho$ as a smooth family of SLDs, which corresponds to setting $K(\rho) = 0$ in the theorem. Remarkably, the following fact holds for $2 \times 2$ matrix representation.

**Corollary 31:** The dualistic structure for $2 \times 2$ matrix representation has vanishing torsion iff it is Riemannian.

**Proof:** For the pure state $\rho$ represented by the following $2 \times 2$ matrix

$$
\rho_{xyz} = \frac{1}{2}(I + x\sigma_x + y\sigma_y + z\sigma_z), \quad x^2 + y^2 + z^2 = 1,
$$

the kernel $K_{sa}(\rho)$ becomes one-dimensional, and is generated by

$$
K = \frac{1}{2}(I - x\sigma_x - y\sigma_y - z\sigma_z).
$$

Letting $K_i = f_i K$, $K_j = f_j K$,

$$
\partial_iK_j - \partial_jK_i = (\partial_if_j)K - (\partial_jf_i)K + f_j(\partial_iK) - f_i(\partial_jK).
$$

The first two terms in the right-hand side belong to $K_{sa}(\rho)$, whereas the other terms become so iff $f_i = f_j = 0$ since both $\partial_iK$ and $\partial_jK$ do not belong to $K_{sa}(\rho)$.

### 3.7. Conclusions

A quantum estimation theory of the pure state models was presented. We first investigated a general framework of the pure state estimation theory and derived quantum counterpart of the Fisher metric. The statistical significance of the Fubini–Study metric was also stressed. We then formulated the one-parameter pure state estimation theory based on the symmetric logarithmic derivative and disclosed the characteristics of the pure states. Some examples were also given in order to demonstrate the one-parameter pure state estimation, and clarified the difference between the pure state models and the strictly positive models.
We next considered the possibility of the estimation theory based on the right logarithmic derivatives. We then investigated $\mathfrak{D}$–invariancy of the SLD–tangent space, which led to the notion of coherent models, and derived explicitly the RLD–bound. Some examples were also given.

We further derived an explicit infimum of $\text{tr} V[M]$ with respect to the linear random measurement $M$ in the linear span of SLDs. The corresponding lower bound was attainable by infinitely many (continuous potency) measurements. The RLD–bound for pure coherent models was also studied, and was found to be most informative.

The construction of the general quantum multi-parameter estimation theory is left for future study, as the strictly positive model case.

Finally, we considered naively the information geometry of quantum pure states. The dualistic structure for pure states was determined by specifying a smooth family of kernels. The curvature is inevitable, whereas the torsion may vanish, whose necessary and sufficient condition was also derived. This attempt has not been completed and most problems are left for future study. For instance, a contrastive study of autoparallelities with respect to various e-connections (including Levi–Civita connection) shall be urgently challenged.

Appendix

A. Parameter Estimation of a Quantum Gaussian Model

In this appendix, we present an accessible review of the parameter estimation theory of quantum Gaussian models. We start with a brief summary of the Weyl representation of Hilbert-Schmidt operators and the non-commutative Fourier transforms under the Fock space representation. For details, see p. 223 [17] or p. 178 [19]. For an arbitrary Hilbert-Schmidt operator $A$, define

$$\mathcal{F}_{x,k}\{A\} = \text{Tr} A \exp \left[ \frac{i}{\hbar} (kQ + xP) \right] = a(x,k), \quad (64)$$

where $[Q,P] = i\hbar$ and $\hbar$ a positive constant (Planck’s constant), then

$$A = \int a(x,k) \exp \left[ -\frac{i}{\hbar} (kQ + xP) \right] \frac{dxdk}{2\pi\hbar} \quad (65)$$

holds. (64) is called the non-commutative Fourier transform of $A$, and (65) is called the Weyl representation of $A$. In particular, the non-commutative
Fourier transform $\mathcal{F}_{x,k}\{\rho\}$ of a density operator $\rho$ is called the quantum characteristic function. With the help of the Baker–Hausdorff formula, we have

$$
\mathcal{F}_{x,k}\{\rho\} = \text{Tr} \, \rho \, \exp \left[ \frac{i}{2\hbar} kx \right] \exp \left[ \frac{i}{\hbar} kQ \right] \exp \left[ \frac{i}{\hbar} xP \right]
$$

$$
= \text{Tr} \, \rho \, \exp \left[ -\frac{i}{2\hbar} kx \right] \exp \left[ \frac{i}{\hbar} xP \right] \exp \left[ \frac{i}{\hbar} kQ \right].
$$

Differentiating these equations, we have useful formulae. For instance,

$$
\frac{\partial}{\partial k} \mathcal{F}_{x,k}\{\rho\} = \text{Tr} \, \rho \left( \frac{i}{2\hbar} x + \frac{i}{\hbar} Q \right) \exp \left[ \frac{i}{2\hbar} kx \right] \exp \left[ \frac{i}{\hbar} kQ \right] \exp \left[ \frac{i}{\hbar} xP \right]
$$

$$
= \frac{i}{2\hbar} x \mathcal{F}_{x,k}\{\rho\} + \frac{i}{\hbar} \mathcal{F}_{x,k}\{\rho Q\},
$$

yields

$$
\mathcal{F}_{x,k}\{\rho Q\} = \left[ -\frac{x}{2} - i\hbar \frac{\partial}{\partial k} \right] \mathcal{F}_{x,k}\{\rho\}.
$$

In the same way,

$$
\mathcal{F}_{x,k}\{Q\rho\} = \left[ \frac{x}{2} - i\hbar \frac{\partial}{\partial k} \right] \mathcal{F}_{x,k}\{\rho\}
$$

$$
\mathcal{F}_{x,k}\{\rho P\} = \left[ \frac{k}{2} - i\hbar \frac{\partial}{\partial x} \right] \mathcal{F}_{x,k}\{\rho\}
$$

$$
\mathcal{F}_{x,k}\{P\rho\} = \left[ -\frac{k}{2} - i\hbar \frac{\partial}{\partial x} \right] \mathcal{F}_{x,k}\{\rho\}.
$$

Then we obtain the following formulae

$$
\mathcal{F}_{x,k}\{\rho Q - Q\rho\} = -x \mathcal{F}_{x,k}\{\rho\}, \quad \mathcal{F}_{x,k}\{\rho Q + Q\rho\} = -2i\hbar \frac{\partial}{\partial k} \mathcal{F}_{x,k}\{\rho\}
$$

$$
\mathcal{F}_{x,k}\{\rho P - P\rho\} = +k \mathcal{F}_{x,k}\{\rho\}, \quad \mathcal{F}_{x,k}\{\rho P + P\rho\} = -2i\hbar \frac{\partial}{\partial x} \mathcal{F}_{x,k}\{\rho\}
$$

The next formula is also useful, which translates the Weyl representation into Glauber-Sudarshan’s P-representation, and vice versa ([6], p. 179 (8.142), p. 185 (8.159) [19]).

**Proposition 32:** The two representations of a trace class operator $T$

$$
T = \int t(x, k) \exp \left[ -\frac{i}{\hbar} (kQ + xP) \right] \frac{dx dk}{2\pi \hbar} = \int \varphi(p, q) |p, q\rangle \langle p, q| \frac{dp dq}{2\pi \hbar}
$$

are related by

$$
\tilde{\varphi}(x, k) = t(x, k) \exp \left[ \frac{1}{4\hbar} (x^2 + k^2) \right],
$$
where \( \hat{\varphi}(x, k) \) is the Fourier transform of \( \varphi(p, q) \), i.e.,

\[
\hat{\varphi}(x, k) = \int \varphi(p, q) \exp \left[ \frac{i}{\hbar} (kq + xp) \right] dpdq.
\]

A Gaussian state of a harmonic oscillator (Boson) is a model for the state of coherent lights associated with thermal noise. A Gaussian state of one degree of freedom with angular frequency \( \omega \) is defined by Glauber-Sudarshan’s P-representation as

\[
\rho_\theta = \frac{1}{\pi} \int \varphi_\theta(z)|z\rangle\langle z| d^2z,
\]

where

\[
\theta = \frac{\omega q + ip}{\sqrt{2 \hbar \omega}}, \quad z = \frac{\omega x + iy}{\sqrt{2 \hbar \omega}}, \quad d^2z = \frac{dxdy}{2\hbar},
\]

with \( q, p \) being the expectation of position and momentum, respectively. Further, \( |z\rangle \) is the Boson coherent state and

\[
\varphi_\theta(z) = \frac{1}{\langle N \rangle} \exp \left[ -\frac{|z - \theta|^2}{\langle N \rangle} \right],
\]

where \( \langle N \rangle \) is a positive quantity which is related to the system temperature (Chap. V [17]). The model can be written also in the form

\[
\rho_\theta = U_\theta \rho_0 U_\theta^*, \quad U_\theta = \exp \left[ \frac{i}{\hbar} (pQ - qP) \right].
\]

With the help of Proposition 32, the corresponding quantum characteristic function becomes

\[
\mathcal{F}_{x,k}\{\rho_\theta\} = \text{Tr} \rho_\theta \exp \left[ \frac{i}{\hbar} (kQ + xP) \right]
\]

\[
= \exp \left[ \frac{i}{\hbar} (kq + xp) - \frac{1}{2\hbar^2} (\sigma_Q^2 k^2 + \sigma_P^2 x^2) \right],
\]

where

\[
\sigma_Q^2 = \hbar \left( \langle N \rangle + \frac{1}{2} \right), \quad \sigma_P^2 = \hbar \left( \langle N \rangle + \frac{1}{2} \right)
\]

are the variances of position \( Q \) and momentum \( P \), respectively, which satisfy

\[
\sigma_Q \sigma_P = \hbar \left( \langle N \rangle + \frac{1}{2} \right).
\]

Let us first calculate the SLDs for the Gaussian model (66). From (69),

\[
\frac{\partial}{\partial x} \mathcal{F}_{x,k}\{\rho\} = \left[ \frac{i}{\hbar} p - \frac{\sigma_P^2}{\hbar^2} x \right] \mathcal{F}_{x,k}\{\rho\}, \quad \frac{\partial}{\partial k} \mathcal{F}_{x,k}\{\rho\} = \left[ \frac{i}{\hbar} q - \frac{\sigma_Q^2}{\hbar^2} k \right] \mathcal{F}_{x,k}\{\rho\}.
\]
Then, by using the formulae obtained above, we have
\[
\mathcal{F}_{x,k}\{\rho Q - Q \rho\} = +\frac{i\hbar}{2\sigma^2_P}\mathcal{F}_{x,k}\{\rho(P - p) + (P - p)\rho\},
\]
\[
\mathcal{F}_{x,k}\{\rho P - P \rho\} = -\frac{i\hbar}{2\sigma^2_Q}\mathcal{F}_{x,k}\{\rho(Q - q) + (Q - q)\rho\}.
\] (71)

By using (68) and the Baker–Hausdorff formula,
\[
\frac{\partial \rho}{\partial q} = -\frac{i}{\hbar}(P \rho - \rho P), \quad \frac{\partial \rho}{\partial p} = \frac{i}{\hbar}(Q \rho - \rho Q).
\]

Taking their non-commutative Fourier transformations and using (71), we get
\[
\mathcal{F}_{x,k}\left\{\frac{\partial \rho}{\partial q}\right\} = -\frac{i}{\hbar}\mathcal{F}_{x,k}\{P \rho - \rho P\} = \frac{1}{2\sigma^2_P}\mathcal{F}_{x,k}\{\rho(Q - q) + (Q - q)\rho\},
\]
\[
\mathcal{F}_{x,k}\left\{\frac{\partial \rho}{\partial p}\right\} = \frac{i}{\hbar}\mathcal{F}_{x,k}\{Q \rho - \rho Q\} = \frac{1}{2\sigma^2_Q}\mathcal{F}_{x,k}\{\rho(P - p) + (P - p)\rho\},
\]
yielding
\[
L^S_q = \frac{1}{\sigma^2_Q}(Q - q), \quad L^S_p = \frac{1}{\sigma^2_P}(P - p). \] (72)

We next calculate the RLDs. The next lemma is the key.

**Lemma 33:**
\[
L^S_j = (I + \frac{i}{2} \mathfrak{D})L^R_j.
\]

**Proof:** By using the definition of the commutation operator (23), we have
\[
(A, B)_\rho = \text{Tr} \rho BA^* = \frac{1}{2}\text{Tr} \rho(BA^* + A^* B) + \frac{1}{2}\text{Tr} \rho(BA^* - A^* B)
\]
\[
= \langle A, (I + \frac{i}{2} \mathfrak{D})B \rangle_\rho.
\]

Now, operating an arbitrary bounded \(X\) to the identity
\[
\frac{\partial \rho_\theta}{\partial \theta^j} = \frac{1}{2}(\rho_\theta L^S_j + L^S_j \rho_\theta) = L^R_j \rho_\theta
\]
from the right, and then taking the trace, we have
\[
\langle [L^S_j, X]_\rho = \langle L^R_j, X \rangle_\rho = \langle (I + \frac{i}{2} \mathfrak{D})L^R_j, X \rangle_\rho.
\]

Since \(X\) is arbitrary, we have the lemma. \(\blacksquare\)
Now, comparing (71) and the definition of $\mathcal{D}$ (23), we have
\[ \mathcal{D}Q = \frac{\hbar}{\sigma_P^2} (P - p), \quad \mathcal{D}P = -\frac{\hbar}{\sigma_Q^2} (Q - q). \] (73)

The RLDs for the Gaussian model are then calculated from (72) and (73) together with Lemma 33 as
\[ L^R_p = \frac{1}{\sigma_P^2 \sigma_Q^2 - \hbar^2/4} \left[ \sigma_Q^2 (P - p) + \frac{i\hbar}{2} (Q - q) \right], \]
\[ L^R_q = \frac{1}{\sigma_P^2 \sigma_Q^2 - \hbar^2/4} \left[ \sigma_P^2 (Q - q) - \frac{i\hbar}{2} (P - p) \right]. \] (74)

The SLD– and the RLD–Fisher informations are readily calculated from (72) and (74) as
\[ J^S = \begin{bmatrix} 1/\sigma_P^2 & 0 \\ 0 & 1/\sigma_Q^2 \end{bmatrix}, \] (75)
\[ J^R = \frac{1}{\sigma_P^2 \sigma_Q^2 - \hbar^2/4} \begin{bmatrix} \sigma_Q^2 & -i\hbar/2 \\ i\hbar/2 & \sigma_P^2 \end{bmatrix}, \] (76)
the corresponding Cramér–Rao inequalities being
\[ V[M] \geq (J^S)^{-1} = \begin{bmatrix} \sigma_P^2 & 0 \\ 0 & \sigma_Q^2 \end{bmatrix}, \] (77)
\[ V[M] \geq (J^R)^{-1} = \begin{bmatrix} \sigma_P^2 & i\hbar/2 \\ -i\hbar/2 & \sigma_Q^2 \end{bmatrix}. \] (78)

Then the CR bounds for SLD (15) and RLD (16) becomes, respectively,
\[ g_P V_P[M] + g_Q V_Q[M] \geq g_P \sigma_P^2 + g_Q \sigma_Q^2 = C^S, \] (79)
\[ g_P V_P[M] + g_Q V_Q[M] \geq g_P \sigma_P^2 + g_Q \sigma_Q^2 + \hbar \sqrt{g_P g_Q} = C^R. \] (80)

Obviously $C^R > C^S$. Moreover, it is shown that $C^R$ is most informative ([32], p. 281 [17]).

B. Spin Coherent State

In this appendix, we give a comprehensible review of spin coherent states. Throughout this appendix, adjoint operators and complex conjugate numbers are denoted by $\dagger$ and $\ast$, respectively, according to the convention in physics. A spin coherent state, originally introduced through a mathematical analogy of the canonical coherent state, permits some fashion of definition since it has been published almost at the same time by several authors.
Here we adopt a rather natural description** of it, based on Radcliffe’s pioneering work [30] and Arecchi et al.’s elegant description [4].

Let $\varepsilon_{abc}$ be the completely anti-symmetric tensor and define the $su(2)$ algebra by

$$[J_a, J_b] = i\varepsilon_{abc} J_c, \quad (a, b, c \in \{x, y, z\}).$$

Further, we take a complete orthonormal basis in the irreducible representation of $su(2)$ with highest weight $j$ as the eigenvectors of $J_z$, i.e.

$$J_z |j, m\rangle = m |j, m\rangle, \quad (m = j, j - 1, \cdots, -j).$$

In the following we write $|j, m\rangle = |m\rangle$ for short.

Now, the spin coherent state $|\mu\rangle$ ($\mu \in \mathbb{C}$) is defined by

$$|\mu\rangle = N^{-1/2} e^{\mu J_ - } |j\rangle = N^{-1/2} \sum_{p=0}^{2j} \binom{2j}{p}^{1/2} \mu^p |j - p\rangle,$$ (81)

where $|j\rangle$ is the highest occupied state, $J_ - = J_x - i J_y$ the spin annihilation operator, $N$ the normalization factor, and the next identity

$$(J_-)^p |j\rangle = \left( \frac{p! 2j!}{(2j - p)!} \right)^{1/2} |j - p\rangle$$ (82)

is used in the second equality. This definition is comparable to that of the canonical coherent state $|z\rangle$, i.e.

$$|z\rangle = e^{-|z|^2/2} e^{za^\dagger} e^{-z^* a} |0\rangle = e^{-|z|^2/2} e^{za^\dagger} |0\rangle.$$ (83)

From (81),

$$\langle \mu | \mu \rangle = N^{-1} \sum_{p=0}^{2j} \binom{2j}{p} |\mu|^2 p = N^{-1} (1 + |\mu|^2)^{2j},$$

then the normalized spin coherent state becomes

$$|\mu\rangle = \frac{1}{(1 + |\mu|^2)^j} e^{\mu J_- } |j\rangle = \frac{1}{(1 + |\mu|^2)^j} \sum_{p=0}^{2j} \binom{2j}{p}^{1/2} \mu^p |j - p\rangle.$$ (83)

We next list up some useful mathematical properties of the spin coherent state, which are straightforward consequences of the representation (83).

**This description was adopted by Lieb in [21].
The inner product is
\[
\langle \lambda | \mu \rangle = \frac{(1 + \lambda^* \mu)^{2j}}{(1 + |\lambda|^2)^j(1 + |\mu|^2)^j}, \tag{84}
\]
\[
|\langle \lambda | \mu \rangle|^2 = \left(1 - \frac{|\lambda - \mu|^2}{(1 + |\lambda|^2)(1 + |\mu|^2)}\right)^{2j}. \tag{85}
\]

The resolution of identity becomes
\[
\frac{2j + 1}{\pi} \int \frac{d^2 \mu}{(1 + |\mu|^2)^2} |\mu\rangle\langle \mu| = \sum_{p=0}^{2j} |j - p\rangle\langle j - p| = 1. \tag{86}
\]

Further, we have
\[
\langle \lambda | J_+ | \mu \rangle = j \frac{2\mu}{1 + \lambda^* \mu} \langle \lambda | \mu \rangle, \quad \langle \lambda | J_- | \mu \rangle = j \frac{2\lambda^*}{1 + \lambda^* \mu} \langle \lambda | \mu \rangle,
\]
\[
\langle \lambda | J_z | \mu \rangle = j \frac{1 - \lambda^* \mu}{1 + \lambda^* \mu} \langle \lambda | \mu \rangle.
\]

The following formula is also useful.
\[
\langle \lambda | J_k^* J_l | \mu \rangle = \frac{1}{(1 + |\lambda|^2)^j(1 + |\mu|^2)^j} \left(\frac{\partial}{\partial \lambda^*}\right)^k \left(\frac{\partial}{\partial \mu}\right)^l (1 + \lambda^* \mu)^{2j}. \tag{87}
\]

Finally, we present an intuitive physical interpretation of spin coherent states. Let \((\theta, \varphi)\) the polar coordinate system on a unit sphere in which the north pole corresponds to \(\theta = 0\) and the \(x\)-axis corresponds to \(\varphi = 0\). The rotation at an angle \(\theta\) with respect to the axis
\[
\mathbf{n} = (\sin \varphi, -\cos \varphi, 0)
\]
is then described by the operator
\[
R[\theta, \varphi] = e^{i\theta (J_x \sin \varphi - J_y \cos \varphi)} = e^{\xi J_- - \xi^* J_+} = e^{\mu J_-} e^{-\log(1 + |\mu|^2) J_z} e^{-\mu^* J_+}. \tag{88}
\]

Here
\[
\xi = \frac{\theta}{2} e^{i\varphi}, \quad \mu = \tan \frac{\theta}{2} e^{i\varphi}, \tag{89}
\]
and the third equality in (88) is called the disentangling theorem [4], which corresponds to the Baker–Hausdorff formula in the Heisenberg algebra. By using (88), the spin coherent state can be rewritten as
\[
|\mu\rangle = (1 + |\mu|^2)^{-j} e^{\mu J_-} |j\rangle = e^{\mu J_-} e^{-\log(1 + |\mu|^2) J_z} e^{-\mu^* J_+} |j\rangle = R[\theta, \varphi]|j\rangle.
\]
This expression indicates that the spin coherent state is given by operating the rotation \(R[\theta, \varphi]\) to the highest occupied state (north pole) \(|j\rangle\). Here \(\mu\)
and \((\theta, \varphi)\) are mutually connected by (89). Geometrically, this correspondence is the stereo projection from the south pole to the tangent space at the north pole. In this sense, the spin coherent state is also denoted as \(|\mu\rangle = |\theta, \varphi\rangle\).

C. Proof of Inequality (7.33)

Denote the right-hand side of (60) by \(f(\theta, \varphi)\), and set \(b = 1\) without loss of generality, i.e.,

\[
f(\theta, \varphi) = \frac{1}{\sin(\theta - \varphi)} \left[ \sqrt{a + \cos 2\theta} + \sqrt{a + \cos 2\varphi} \right], \quad (a > 1). \tag{90}
\]

Extremizing conditions \(\partial f / \partial \theta = 0\) and \(\partial f / \partial \varphi = 0\) lead to

\[
\left[ a + \cos 2\theta + \sqrt{(a + b \cos 2\theta)(a + b \cos 2\varphi)} \right] \cot(\theta - \varphi) + \sin 2\theta = 0, \tag{91}
\]

and

\[
\left[ a + \cos 2\varphi + \sqrt{(a + b \cos 2\theta)(a + b \cos 2\varphi)} \right] \cot(\theta - \varphi) - \sin 2\varphi = 0, \tag{92}
\]

respectively. Subtracting these two equalities, we have

\[
(\cos 2\theta - \cos 2\varphi) \cot(\theta - \varphi) + \sin 2\theta + \sin 2\varphi = 0,
\]

but this is an identity. So the equalities (91) and (92) are equivalent. In other words, (91) is the only extremizing condition for (90). On the other hand, adding the equalities (91) and (92), we have,

\[
\cos(\theta - \varphi) = -\frac{2}{f^2(\theta, \varphi)} \cos(\theta + \varphi), \tag{93}
\]

which is also equivalent to either (91) or (92). We show that, under the condition \(0 < \theta - \varphi < \pi\), (93) have a unique globally connected solution. Suppose (93) have possibly disconnected solutions (curves) labeled by \(C_n\). Note that \(f(\theta, \varphi)\) is constant on each \(C_n\) since \(\partial f / \partial \theta = \partial f / \partial \varphi = 0\) along the curve. Let us denote the corresponding constant value of \(2/f^2(\theta, \varphi)\) on \(C_n\) by \(A_n\). Furthermore, let us change the coordinate system as \(\theta - \varphi = \pi/2 - \delta, \theta + \varphi = \pi/2 - x\). Then

\[
f(\theta, \varphi) = \hat{f}(x, \delta) = \frac{1}{\cos \delta} \left[ \sqrt{a + \cos(x - \delta)} + \sqrt{a - \cos(x + \delta)} \right], \quad (a > 1). \tag{94}
\]

and the extremizing condition (93) becomes

\[
\sin \delta = -A_n \sin x, \quad (-\frac{\pi}{2} < \delta < \frac{\pi}{2}). \tag{95}
\]
Therefore, we can regard $C_n$ as the connected component of the solution of (95) which crosses the $x$-axis at $x = n\pi$ ($n = 0, \pm1, \cdots$). Since $A_n$ is constant on each curve $C_n$, we can evaluate its value at the $x$-intercepts as

$$A_n = \frac{2}{\hat{f}^2(n\pi, 0)} = \frac{2}{(\sqrt{a+1} + \sqrt{a-1})^2}.$$  

In particular, $A_n$ does not depend on $n$ and is less than unity. Therefore, the extremizing condition (95), which is rewritten as

$$\sin \delta = -\frac{2}{(\sqrt{a+1} + \sqrt{a-1})^2} \sin x,$$

has a unique globally connected solution

$$\delta = -\arcsin \left[ \frac{2}{(\sqrt{a+1} + \sqrt{a-1})^2} \sin x \right]. \quad (96)$$

In other words, all the $C_n$’s are identical with each other. It is evident that $\hat{f}(x, \delta)$ takes its minimum along the curve (96). Therefore,

$$f(\theta, \varphi) \geq \hat{f}(0, 0) = \sqrt{a+1} + \sqrt{a-1}.$$ 

D. Photon Squeezed State

In this appendix, we give a quick review of photon squeezed states. Throughout this appendix, adjoint operators and complex conjugate numbers are denoted by $\dagger$ and $\ast$, respectively, according to the convention in physics. Let $a, a^\dagger$ be the Boson annihilation and creation operators, respectively, which satisfy $[a, a^\dagger] = 1$.

**Proposition 34:** The following relations hold.

(i) $[a^2, a] = -2a^\dagger$, $[a^2, a^\dagger] = 2a$.

(ii) Given $\xi \in \mathbb{C}$, define $A = \frac{1}{2}(\xi a^\dagger^2 - \xi^* a^2)$. Then it satisfies $A^\dagger = -A$, $[A, a] = -\xi a^\dagger$, $[A, a^\dagger] = -\xi^* a$.

(iii) Define $b = e^{-A} a e^A$, then $b^\dagger = e^{-A} a^\dagger e^A$, $[b, b^\dagger] = 1$ and

$$b = a \cosh |\xi| + a^\dagger \frac{\xi}{|\xi|} \sinh |\xi|, \quad a = b \cosh |\xi| - b^\dagger \frac{\xi}{|\xi|} \sinh |\xi|.$$ 

(iv) $e^{-A} e^{-z^a^\dagger - z^* a} e^A = e^{-b^\dagger - z^* b}$.

(v) $za^\dagger - z^* a = \beta b^\dagger - \beta^* b$ where $\beta = z \cosh |\xi| + z^* \frac{\xi}{|\xi|} \sinh |\xi|$. 
**Proposition 38:** These fashions are related by the following proposition.

**(i)** is immediately derived from (i). Further, (iii) is derived as

\[
e^{-A}a e^A = a + [-A, a] + \frac{1}{2!}[-A, [-A, a]] + \cdots
\]

= \(a + \xi a^\dagger + \frac{1}{2!} \xi^2 a + \frac{1}{3!} \xi^2 \xi a^\dagger + \cdots\)

= \(a \left(1 + \frac{\xi^2}{2!} + \cdots\right) + a^\dagger \frac{\xi}{|\xi|} \left(|\xi| + \frac{\xi^3}{3!} + \cdots\right) = a \cosh |\xi| + a^\dagger \frac{\xi}{|\xi|} \sinh |\xi|\).

(iv) is obtained by a Taylor expansion of \(e^{za^\dagger - z^* a}\) and using (iii). (v) is a straightforward consequence of (iii).

Let us define the following unitary operators.

\(D(z) = \exp(za^\dagger - z^* a), \quad S(\xi) = \exp\left[\frac{1}{2} (\xi a^\dagger a - \xi^* a^2)\right], \quad \hat{D}_\xi(z) = \exp(zb^\dagger - z^* b)\).

\(D(z)\) and \(S(\xi)\) are called, respectively, the shift operator and the squeezing operator. Clearly \(D^\dagger(z) = D^{-1}(z) = D(-z)\) etc. hold and the next formulae are derived immediately from the above relations (iii) (iv) and (v).

**Proposition 35:** The following relations hold.

(i) \(b = S^{-1}(\xi) a S(\xi) = a \cosh |\xi| + a^\dagger \frac{\xi}{|\xi|} \sinh |\xi|\).

(ii) \(S^{-1}(\xi) D(z) S(\xi) = \hat{D}_\xi(z)\).

(iii) \(D(z) = \hat{D}_\xi(\beta), \quad \beta = z \cosh |\xi| + z^* \frac{\xi}{|\xi|} \sinh |\xi|\).

Now let us define the squeezed state. There are mainly two fashions to define the photon squeezed state [5, 33].

**Definition 36:** (Yuen) \(b|\beta\rangle_\xi = |\beta\rangle_\xi\).

**Definition 37:** (Caves) \(|z\rangle_\xi = D(z) S(\xi)|0\rangle\).

These fashions are related by the following proposition.

**Proposition 38:** The following relations hold.

(i) \(|\beta\rangle_\xi = S^{-1}(\xi)|\beta\rangle\).

(ii) \(|\beta\rangle_\xi = \hat{D}_\xi(\beta)|0\rangle_\xi\).

(iii) \(|\beta\rangle_\xi = |z\rangle_{(-\xi)}\).

Proof:

(i) \(b S^{-1}(\xi) |\beta\rangle = \{S^{-1}(\xi) a S(\xi)\} S^{-1}(\xi) |\beta\rangle = \beta S^{-1}(\xi) |\beta\rangle\).

(ii) \(|\beta\rangle_\xi = S^{-1}(\xi) |\beta\rangle = S^{-1}(\xi) D(\beta)|0\rangle = \hat{D}_\xi(\beta) S^{-1}(\xi) |0\rangle = \hat{D}_\xi(\beta)|0\rangle_\xi\).
We can use both expressions of a squeezed state interchangeably through the formula (iii) in Proposition 38, provided that when we write as $|z\rangle_\xi = |\beta\rangle_{(-\xi)}$, all the $\xi$ in $b$ and $\beta$ must automatically be changed into $-\xi$ \textsuperscript{††}.

Now, let us calculate some expectational properties of squeezed states. Let the real and imaginary parts of $a$ be $X_r$, $X_i$, respectively, and the real and imaginary parts of $z$ be $x_r$, $x_i$, respectively. Introducing

$$\mu = \cosh |\xi|, \quad \nu = \frac{\xi}{|\xi|} \sinh |\xi|,$$

which satisfy

$$|\mu|^2 - |\nu|^2 = 1, \quad b = \mu a + \nu a^\dagger, \quad \beta = \mu z + \nu z^*, \quad a = \mu^* b - \nu b^\dagger,$$

the expectation of $X_r$ can be calculated as

$$\langle\beta|X_r|\beta\rangle = \langle\beta| \frac{a + a^\dagger}{2} |\beta\rangle = \frac{1}{2} \langle\beta| (\mu - \nu)^* b + (\mu - \nu) b^\dagger |\beta\rangle$$

$$= \frac{1}{2} [(\mu - \nu)^* \beta + (\mu - \nu) \beta^*] = \frac{1}{2} (z + z^*) (|\mu|^2 - |\nu|^2) = x_r.$$

In the same way, we have

$$\langle\beta|X_i|\beta\rangle = x_i.$$

Further, by using $[b, b^\dagger] = 1$,

$$\langle\beta|(X_r - x_r)^2 |\beta\rangle = \langle\beta|X_r^2|\beta\rangle - \langle\beta|X_r|\beta\rangle^2$$

$$= \langle\beta| \left( \frac{a + a^\dagger}{2} \right)^2 |\beta\rangle - \langle\beta| \frac{a + a^\dagger}{2} |\beta\rangle^2$$

$$= \frac{1}{4} \left[ \langle\beta|[ (\mu - \nu)^* b^2 + (\mu - \nu)^* (\mu - \nu) (b b^\dagger + b^\dagger b) + (\mu - \nu)^2 b^\dagger b^2 |\beta\rangle \right]$$

$$\quad - \frac{1}{4} [(\mu - \nu)^* \beta + (\mu - \nu) \beta^*]^2$$

$$= \frac{1}{4} [ (\mu - \nu)^* \beta^2 + |\mu - \nu|^2 (1 + 2 |\beta|^2) + (\mu - \nu)^2 \beta^*^2 ]$$

$$\quad - \frac{1}{4} [(\mu - \nu)^* \beta + (\mu - \nu) \beta^*]^2$$

$$= \frac{1}{4} |\mu - \nu|^2.$$

\textsuperscript{††}As we shall see later, $|z\rangle_\xi$ and $|z\rangle_{(-\xi)}$ are states which have the same center and principal axes but have opposite squeezing ratios.
Thus \( V^\dagger \) and \( V \) respectively, which is illustrated in Figure 3. These results indicate that the covariances of a squeezed state are independent of the complex amplitude \( z \).

Expressing \( \xi = se^{i\theta} \) and \( \mu = \cosh s \), \( \nu = -e^{i\theta} \sinh s \), the covariance matrix of \( X_r, X_i \) for the state \( |z\rangle \xi = |\beta\rangle \langle -\xi| \) becomes

\[
V = \frac{1}{4} \begin{bmatrix}
    \cosh 2s + \cos \theta \sinh 2s & \sin \theta \sinh 2s \\
    \sin \theta \sinh 2s & \cosh 2s - \cos \theta \sinh 2s
\end{bmatrix}
\]

\[
= \frac{1}{4} \begin{bmatrix}
    e^{2s} \cos^2 \frac{\theta}{2} + e^{-2s} \sin^2 \frac{\theta}{2} & (e^{2s} - e^{-2s}) \sin \theta \cos \frac{\theta}{2} \\
    (e^{2s} - e^{-2s}) \sin \theta \cos \frac{\theta}{2} & e^{2s} \sin^2 \frac{\theta}{2} + e^{-2s} \cos^2 \frac{\theta}{2}
\end{bmatrix}
\]

\[
= \frac{1}{4} O \left( \frac{\theta}{2} \right) \begin{bmatrix}
    e^{2s} & 0 \\
    0 & e^{-2s}
\end{bmatrix} O^{-1} \left( \frac{\theta}{2} \right), \tag{97}
\]

where, \( O(\varphi) \) is the orthogonal (rotational) matrix defined by

\[
O(\varphi) = \begin{bmatrix}
    \cos \varphi - \sin \varphi \\
    \sin \varphi & \cos \varphi
\end{bmatrix}.
\]

Thus \( V \) has eigenvalues \( e^{2s}/4 \) and \( e^{-2s}/4 \), and the corresponding eigenvectors are

\[
\begin{bmatrix}
    \cos \frac{\theta}{2} \\
    \sin \frac{\theta}{2}
\end{bmatrix}, \quad \begin{bmatrix}
    -\sin \frac{\theta}{2} \\
    \cos \frac{\theta}{2}
\end{bmatrix},
\]

respectively, which is illustrated in Figure 3. Moreover, since the corresponding principal axes \( u \) and \( v \) are related to the original coordinate system \((x, y)\) by

\[
\begin{bmatrix}
    x \\
    y
\end{bmatrix} = \begin{bmatrix}
    \cos \frac{\theta}{2} - \sin \frac{\theta}{2} \\
    \sin \frac{\theta}{2} & \cos \frac{\theta}{2}
\end{bmatrix} \begin{bmatrix}
    u \\
    v
\end{bmatrix},
\]

\[\text{Since} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \iff (x y) \begin{bmatrix}
    1/a^2 & 0 \\
    0 & 1/b^2
\end{bmatrix} \begin{bmatrix}
    x \\
    y
\end{bmatrix} = 1, \quad \text{principal lengths of an ellipsoid are the squared roots of the inverse of the eigenvalues of the corresponding quadratic form. On the other hand, the inverse of covariance matrix determines a quadratic form, as is exemplified by the Gaussian distribution. Thus the squared roots of the eigenvalues of covariance matrix gives the principal lengths.\]
and since the equation of the ellipse centered at the origin becomes

\[ u = \frac{1}{2} e^s \cos \phi, \quad v = \frac{1}{2} e^{-s} \sin \phi, \]

we have

\[ x = \frac{e^s}{2} \cos \frac{\theta}{2} \cos \phi - \frac{e^{-s}}{2} \sin \frac{\theta}{2} \sin \phi = \sqrt{\frac{e^{2s}}{4} \cos^2 \frac{\theta}{2} + \frac{e^{-2s}}{4} \sin^2 \frac{\theta}{2} \cos (\phi + \alpha)}, \]

\[ y = \frac{e^s}{2} \sin \frac{\theta}{2} \cos \phi + \frac{e^{-s}}{2} \cos \frac{\theta}{2} \sin \phi = \sqrt{\frac{e^{2s}}{4} \sin^2 \frac{\theta}{2} + \frac{e^{-2s}}{4} \cos^2 \frac{\theta}{2} \cos (\phi + \beta)}. \]

These equations indicate that the widths of orthogonal projections of the ellipse onto the \(x, y\) axes are identical to the variances of \(X_r, X_i\), respectively*.

References


*The covariance matrix of \(Q, P\) is \(2V\), since \(X_r + iX_i = (Q + iP)/\sqrt{2}\).


H. Nagaoka, unpublished manuscript‡.


CHAPTER 19

An Estimation Theoretical Characterization of Coherent States

Akio Fujiwara and Hiroshi Nagaoka

Abstract. We introduce a class of quantum pure state models called the coherent models. A coherent model is an even dimensional manifold of pure states whose tangent space is characterized by a symplectic structure. In a rigorous framework of noncommutative statistics, it is shown that a coherent model inherits and expands the original spirit of the minimum uncertainty property of coherent states.

1. Introduction

Quantum estimation theory, originated in optical communications, offers a rigorous approach toward the optimization of detection processes in quantum communication systems [7, 8]. It aims to find, for a given smooth parametric family of density operators (a model) $\mathcal{P} = \{\rho_\theta; \theta = (\theta^1, \ldots, \theta^n) \in \Theta \subset \mathbb{R}^n\}$, the optimum measurement (positive operator-valued measure) $M = \{M(B); B$ is a Borel set in $\mathbb{R}^n\}$ for the parameter $\theta$ under the unbiasedness condition: For all $\theta \in \Theta$,

$$\int \hat{\theta}^j \text{Tr} \rho_\theta M(d\hat{\theta}) = \theta^j, \quad j = 1, \ldots, n.$$ 

Here $\text{Tr}$ denotes the operator trace. Normally a more tractable (weaker) condition is adopted, called the local unbiasedness condition: A measurement $M$ is called locally unbiased at a given point $\theta$ if $M$ satisfies at $\theta$ the above equality and its formal differentiation

$$\frac{\partial}{\partial \theta^i} \int \hat{\theta}^j \text{Tr} \rho_\theta M(d\hat{\theta}) = \delta_i^j, \quad i, j = 1, \ldots, n.$$ 

It is well-known that when $n = 1$, the quantum Cramér-Rao inequality with respect to the symmetric logarithmic derivative (SLD) offers the achievable lower bound (i.e., the bound attained by a certain measurement) of the variance of estimation. This is also regarded as a rigorous modification

of the uncertainty relation. When \( n \geq 2 \), on the other hand, a matrix version of the SLD Cramér-Rao inequality itself does not always have an absolute significance because the lower bound cannot be attained in general unless the model has commutative SLDs. We therefore often deal with the minimization problem of the scalar quantity \( \text{tr} GV_\theta[M] \) with respect to \( M \), where \( \text{tr} \) denotes the matrix trace on the parameter space \( \Theta \), \( G \) a real symmetric positive matrix representing the weight, and \( V_\theta[M] \) the covariance matrix at \( \theta \) with respect to a locally unbiased measurement \( M \) whose \((i,j)\) entry is

\[
(V_\theta[M])^{ij} = \int (\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j) \text{Tr} \rho_\theta M(d\hat{\theta}).
\]

If there is a number \( C \) such that \( \text{tr} GV_\theta[M] \geq C \) holds for all \( M \), \( C \) is called a Cramér-Rao type bound or simply a CR bound. The CR bound \( C \) may depend on both \( G \) and \( \theta \). The problem of finding the achievable CR bound is in general a hard one and has been solved only in a few special models such as the quantum gaussian model [16, 8] and the two-dimensional spin 1/2 model [12, 13].

Holevo showed that if a model having the right logarithmic derivative (RLD) exhibits a certain “nice” property of a tangent space, the CR bound based on the RLD is expressed only in terms of the SLDs (p.280 [8]). Moreover it was shown that this gives the achievable CR bound for the gaussian model of quantum oscillators. Motivated by these facts and that the SLD Fisher information is well-defined also for pure state models [4], we will introduce a class of pure state models called the coherent models [5] each having a “nice” tangent space, and will explore their parameter estimation theory.

The construction of the paper is as follows. In Section 2, we explore some basic characteristics inherent in pure state models which are closely related with Holevo’s commutation operator. In Section 3, a special class of pure state models, called the coherent models, is introduced of which the SLD tangent space forms an invariant subspace with respect to the commutation operator. In Section 4, we derive a CR bound, called the generalized RLD bound, for a model that has an invariant SLD tangent space with respect to the commutation operator. Here the model is not assumed to be pure. In Section 5, we show that for a coherent model, there exists a random measurement which attains the generalized RLD bound. In Section 6, the above results are demonstrated in two simple coherent models: a canonical squeezed state model and a spin coherent state model. In the final section
we give conclusions.

2. Commutation Operator

In the study of noncommutative statistics, Holevo introduced useful mathematical tools called the square summable operators and the commutation operators associated with quantum states. We here give a brief summary: for details, consult [8]. Let $\mathcal{H}$ be a separable complex Hilbert space which corresponds to a physical system of interest, and let $\rho$ be a fixed density operator. We define a real Hilbert space $L^2_h(\rho)$ associated with $\rho$ by the completion of $B_h(\mathcal{H})$, the set of bounded self-adjoint operators, with respect to the pre-inner product $\langle X,Y \rangle_\rho = \text{Re} \text{ Tr} \rho XY$. Letting $\rho = \sum_j s_j |\psi_j \rangle \langle \psi_j |$ be the spectral representation, an element $X \in L^2_h(\rho)$ can be regarded as an equivalence class of such self-adjoint operators (called square summable operators) satisfying $\sum_j s_j \|X\psi_j\|^2 < \infty$ (so that $\psi_j \in \text{Dom}(X)$ if $s_j \neq 0$) under the identification $X_1 \sim X_2$ if $X_1\psi_j = X_2\psi_j$ for $s_j \neq 0$. The space $L^2_h(\rho)$ thus provides a convenient tool to cope with unbounded observables. Let $L^2(\rho)$ be the complexification of $L^2_h(\rho)$. Note that $L^2(\rho)$ is also regarded as the completion of $B(\mathcal{H})$, the set of bounded operators, with respect to the pre-inner product

$$\langle X,Y \rangle_\rho = \frac{1}{2} \text{ Tr} \rho (YX^* + X^*Y).$$

Thus $L^2(\rho)$ is regarded as a complex Hilbert space with the inner product $\langle \cdot, \cdot \rangle_\rho$. We further introduce two sesquilinear forms on $B(\mathcal{H})$ by

$$(X,Y)_\rho = \text{ Tr} \rho YX^*, \quad [X,Y]_\rho = \frac{1}{2i} \text{ Tr} \rho (YX^* - X^*Y),$$

and extend them to $L^2(\rho)$ by continuity.

The commutation operator $D_\rho : L^2(\rho) \to L^2(\rho)$ with respect to $\rho$ is defined by $[X,Y]_\rho = \langle X,D_\rho Y \rangle_\rho$, which is formally represented by the operator equation $\rho(D_\rho X) + (D_\rho X)\rho = \frac{1}{i} (\rho X - X \rho)$. (To be precise, this definition is different from Holevo's original definition by a factor of 2.) The operator $D_\rho$ is a complex-linear bounded skew-adjoint operator. Moreover, since the forms $\langle \cdot, \cdot \rangle_\rho$ and $\langle \cdot, \cdot \rangle_\rho$ are real on the real subspace $L^2_h(\rho)$, this subspace is invariant under the operation of $D_\rho$. Thus $D_\rho$ can also be regarded as a real-linear bounded skew-adjoint operator when restricted to $L^2_h(\rho)$ as $D_\rho : L^2_h(\rho) \to L^2_h(\rho)$.

Our main concern lies in the case where $\rho$ is pure. In this case the above setting is considerably simplified as follows: Let $\rho = |\psi \rangle \langle \psi |$. Then for
$X, Y \in \mathcal{L}^2(\rho),$

$$\langle X, Y \rangle_\rho = \frac{1}{2} \{\langle Y^*|X^*\psi \rangle + \langle X\psi|Y\psi \rangle\},$$

$$[X, Y]_\rho = \frac{1}{2i} \{\langle Y^*|X^*\psi \rangle - \langle X\psi|Y\psi \rangle\},$$

$$(X, Y)_\rho = \langle Y^*|X^*\psi \rangle.$$

Here $X\psi$, for example, stands for the vector $X_1\psi$ where $X_1$ is an arbitrary representative of $X$. (It is independent of the choice of a representative.) In particular, if $X, Y \in \mathcal{L}^2_\hbar(\rho)$ we have

$$\langle X, Y \rangle_\rho = \operatorname{Re} \langle Y\psi|X\psi \rangle = \operatorname{Re} \langle X\psi|Y\psi \rangle,$$  \hspace{1cm} (1)

$$[X, Y]_\rho = \operatorname{Im} \langle Y\psi|X\psi \rangle = -\operatorname{Im} \langle X\psi|Y\psi \rangle,$$ \hspace{1cm} (2)

$$(X, Y)_\rho = \langle Y\psi|X\psi \rangle = \langle X\psi|Y\psi \rangle.$$ \hspace{1cm} (3)

It should be noted that operators $X$ and $Y$ (whether bounded or not) are identified with each other in $\mathcal{L}^2(\rho)$ iff $X\psi = Y\psi$ and $X^*\psi = Y^*\psi$. In particular, self-adjoint operators $X$ and $Y$ are identified in $\mathcal{L}^2_\hbar(\rho)$ iff $X\psi = Y\psi$.

**Lemma 1:** Let $\rho = |\psi\rangle \langle \psi|$. Then for all $X \in \mathcal{L}^2_\hbar(\rho),$

$$(D_\rho X)\psi = i(X - \langle \psi|X\psi \rangle I)\psi,$$

where $I$ denotes the identity in $\mathcal{L}^2_\hbar(\rho)$.

**Proof:** For $X \in \mathcal{L}^2_\hbar(\rho)$, let $Z$ be the element in $\mathcal{L}^2_\hbar(\rho)$ having a representative $Z_1 = i(|X\psi\rangle \langle \psi| - |\psi\rangle \langle X\psi|)$. Then $Z\psi = i(X - \langle \psi|X\psi \rangle I)\psi$. On the other hand, for $Y \in \mathcal{L}^2_\hbar(\rho)$, we have

$$\langle Y\psi|Z\psi \rangle = i\{\langle Y\psi|X\psi \rangle - \langle \psi|X\psi \rangle \langle \psi|Y\psi \rangle\},$$

and hence $\langle Y, Z \rangle_\rho = [Y, X]_\rho$ because of (1) and (2). Thus $Z = D_\rho X$, which completes the proof. \[\square\]

Note that Lemma 1 does not imply $D_\rho X = i(X - \langle \psi|X\psi \rangle I)$, since the right hand side is not a self-adjoint element in $\mathcal{L}^2(\rho)$ unless it equals 0.

Let us introduce a linear subspace

$$\mathcal{T}_\hbar(\rho) = \{X \in \mathcal{L}^2_\hbar(\rho) : \langle I, X \rangle_\rho = 0\}$$

of $\mathcal{L}^2_\hbar(\rho)$. Here $\rho$ is not necessarily pure. This subspace is itself a real Hilbert space with the inner product $\langle \cdot, \cdot \rangle_\rho$. Now consider again the special case
that $\rho$ is pure: $\rho = |\psi\rangle\langle\psi|$. Then from Lemma 1, we obtain the important relation:

$$(D_\rho X)\psi = (iX)\psi, \quad X \in T_h(\rho).$$

(4)

This equation, combined with (1), implies that $D_\rho$ is a unitary transformation on $(T_h(\rho), \langle\cdot,\cdot\rangle_\rho)$. In particular, $D_\rho$ is nondegenerate on $T_h(\rho)$, and so is the skew-symmetric bilinear form $[\cdot,\cdot]_\rho$. In other words, the real linear space $T_h(\rho)$ is regarded as a symplectic space [6] with the symplectic form $[\cdot,\cdot]_\rho$. We also note that $D_\rho^2 = -I$ on $T_h(\rho)$ ($I$ denotes the identity operator acting on $T_h(\rho)$), since $D_\rho$ is unitary and skew-adjoint. Indeed, equation (4) immediately leads to $(D_\rho^2 X)\psi = -X\psi$ and hence $D_\rho^2 X = -X$ for all $X \in T_h(\rho)$, whereas $D_\rho X \neq iX$ as mentioned earlier. In other words, $D_\rho$ is an almost complex structure on $T_h(\rho)$.

### 3. Coherent Model

Let $\mathcal{P} = \{\rho_\theta ; \theta = (\theta^1, \ldots, \theta^n) \in \Theta\}$ be an $n$-dimensional model, where $\rho_\theta$ are not necessarily pure for the present, and $\Theta$ is an open subset of $\mathbb{R}^n$. We assume the following regularity conditions:

(a) The parametrization $\theta \mapsto \rho_\theta$ is assumed to be appropriately smooth and nondegenerate so that the derivatives $\{\partial \rho_\theta / \partial \theta^j\}_{j=1}^n$ exist in trace-class and form a linearly independent set at each point $\theta$.

(b) There exists a constant $c$ such that

$$\left|\frac{\partial}{\partial \theta^j} \text{Tr} \rho_\theta X\right|^2 \leq c \langle X, X \rangle_{\rho_\theta}$$

for all $X \in \mathcal{B}(\mathcal{H})$ and $j$.

From the condition (b), the linear functionals $X \mapsto (\partial/\partial \theta^j)\text{Tr} \rho_\theta X$ can be extended to continuous linear functionals on $\mathcal{L}^2(\rho_\theta)$.

Given a model that satisfies (a) and (b), the symmetric logarithmic derivative (SLD) $L^S_{\theta,j}$ in the $j$th direction is defined by the requirement that

$$\frac{\partial}{\partial \theta^j} \text{Tr} \rho_\theta X = \langle L^S_{\theta,j}, X \rangle_{\rho_\theta}, \quad L^S_{\theta,j} \in \mathcal{L}^2(\rho_\theta)$$

for all $X \in \mathcal{L}^2(\rho_\theta)$. It is easily verified that $L^S_{\theta,j} \in \mathcal{L}^2_\rho(\rho_\theta)$; so the definition is formally written as $\partial \rho_\theta / \partial \theta^j = \frac{1}{2} L^S_{\theta,j} (\rho_\theta + \rho_\theta L^S_{\theta,j})$. The SLDs belong to $\mathcal{T}_h(\rho_\theta)$ since $\langle I, L^S_{\theta,j} \rangle_{\rho_\theta} = (\partial/\partial \theta^j)\text{Tr} \rho_\theta = 0$, and the SLD Fisher information matrix defined by $J^S_\theta = \left[\langle L^S_{\theta,j}, L^S_{\theta,k} \rangle_{\rho_\theta}\right]$ gives a Cramér-Rao inequality.
$V_0[M] \geq (J_0^S)^{-1}$, where $M$ is an arbitrary locally unbiased measurement for the parameter $\theta$, see p. 276 [8].

In the rest of this section, we restrict ourselves to pure state models. Some remarks are in order. First, by differentiating the identity $\rho_0^2 = \rho_0$, we see that the element in $\mathcal{L}_h^2(\rho_0)$ having a representative $2\partial\rho_0/\partial\theta^j$ gives the SLD $L_{\theta,j}^S$. Thus, for a pure state model, the condition (a) implies (b). Second, associated with a pure state model $\{\rho_\theta; \theta \in \Theta\}$ is, at least locally, a smooth family $\{\psi_\theta; \theta \in \Theta\}$ of normalized vectors in $\mathcal{H}$ such that $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$. In what follows, we shall frequently use this representation.

A convenient way of finding SLDs for a pure state model $\rho_\theta$ is as follows: Let $L_{\theta,j}^A$ be the anti-symmetric logarithmic derivative (ALD) satisfying
\[
\frac{\partial}{\partial\theta^j}\text{Tr} \rho_\theta X = [L_{\theta,j}^A, X]_{\rho_\theta}, \quad L_{\theta,j}^A \in T_h(\rho_\theta)
\]
for all $X \in \mathcal{L}_h^2(\rho_\theta)$, or formally $\partial\rho_\theta/\partial\theta^j = (L_{\theta,j}^A\rho_\theta - \rho_\theta L_{\theta,j}^A)/2i$. (This definition is different from [4] by a factor of $i$.) Then the SLD is given by $L_{\theta,j}^S = -D_\theta L_{\theta,j}^A$ where $D_\theta = D_{\rho_\theta}$, since $(L_{\theta,j}^S, X)_{\rho_\theta} = [L_{\theta,j}^A, X]_{\rho_\theta}$. Note that since $D_\theta^2 = -I$ on $T_h(\rho_\theta)$, then $L_{\theta,j}^A = D_\theta L_{\theta,j}^S$, which assures the existence and the uniqueness of the ALD for a pure state model. The advantage of the use of the ALD is this: Every pure state model can be expressed in the form $\rho_\theta = U_\theta \rho_0 U_\theta^*$ where $\{U_\theta\}_\theta$ is a smooth family of unitary operators (which do not necessarily form a group representation), so that the ALD is explicitly given by
\[
L_{\theta,j}^A = 2i(A_{\theta,j} - \langle I, A_{\theta,j}\rangle_{\rho_\theta}),
\]
where $A_{\theta,j}$ is the skew-adjoint element in $\mathcal{L}_h^2(\rho_\theta)$ having a representative $(\partial U_\theta/\partial\theta^j)U_\theta^*$, the local generator of $U_\theta$. For a group covariant pure state model, the generator of the group is usually obvious.

Let $T_\theta^S(\mathcal{P})$ be the real-linear subspace of $T_h(\rho_\theta)$ spanned by the SLDs $\{L_{\theta,j}^S\}_j$. Since the tangent vectors of the manifold $\mathcal{P}$ at the point $\rho_\theta$ are faithfully represented by the elements of $T_\theta^S(\mathcal{P})$ via the correspondence $(\partial/\partial\theta^j)_\theta \mapsto L_{\theta,j}^S$, we call $T_\theta^S(\mathcal{P})$ the SLD tangent space of the model $\mathcal{P}$ at $\theta$. A pure state model $\mathcal{P} = \{\rho_\theta; \theta \in \Theta\}$ is called locally coherent at $\theta$ if $T_\theta^S(\mathcal{P})$ is $D_\theta$-invariant. The model is called coherent if it is locally coherent for all $\theta \in \Theta$.

When the Hilbert space $\mathcal{H}$ is finite-dimensional, the totality of pure states forms a complex projective space and is an example of coherent model. The Riemannian metric on the model induced by the SLD Fisher information matrix $J_\theta^S$ is identical to the Fubini-Study metric up to a con-
constant factor [4] and hence is a Kähler metric. The associated fundamental 2-form [10] in this case is nothing but the symplectic structure $[\cdot, \cdot]_\rho$.

**Theorem 2:** Consider a pure state model of the form $\rho_\theta = U_{g(\theta)} \rho_0 U_{g(\theta)}^*$ where $\{U_g ; g \in G\}$ is a projective unitary representation of a Lie group $G$ and $g(\cdot) : \theta \mapsto g(\theta)$ is the parametrization of the elements of $G$ by a local coordinate system satisfying $g(0) = e$ (the unit element). This model is coherent iff it is locally coherent at $\rho_0$.

**Proof:** We only need to prove the if part. Let $\Lambda_\theta : G \rightarrow G$ be the left translation by $g(\theta)^{-1}$ which maps $h \mapsto g(\theta)^{-1}h$. Then its differential $(d\Lambda_\theta)_{g(\theta)} : T_{g(\theta)}(G) \rightarrow T_{g(\theta)}(G)$ is represented by a nonsingular real matrix $a^k_j(\theta)$ such that $(d\Lambda_\theta)_{g(\theta)}([\partial g(\theta)/\partial \theta^i]_\theta) = \sum_k a^k_j(\theta) \partial g(\theta)/\partial \theta^k|_{\theta=0}$. Now since $\rho_{\theta+\Delta \theta} = U_{g(\theta)} \rho_{\Delta \theta} U_{g(\theta)}^*$, where $\Lambda_\theta(g(\theta + \Delta \theta)) = g(\Delta \theta')$, we find that

$$\partial \rho_\theta/\partial \theta^j = \sum_k a^k_j(\theta) U_\theta [\partial \rho_\theta/\partial \theta^k]|_{\theta=0} U_\theta^*.$$ 

This implies that the SLDs at $\theta$ are given by $L^S_\theta,j = \sum_i a^k_j(\theta) U_\theta L^S_0,k U_\theta^*$. As a consequence

$$L^S_{\theta,j} \psi_\theta = \sum_k a^k_j(\theta) U_\theta L^S_0,k \psi_0.$$  

(5)

Here we have set as $\rho_\theta = |\psi_\theta\rangle \langle \psi_\theta|$ with $\psi_\theta = U_\theta \psi_0$. Now suppose $\mathcal{P}$ is locally coherent at $\rho_0$. Then the vector $(\mathcal{D}_0 L^S_{0,k}) \psi_0 = i L^S_{0,k} \psi_0$ (see 4) belongs to the real linear span of $\{L^S_{0,k'} \psi_0\}_{k'=1}^n$; hence the vector $(\mathcal{D}_0 L^S_{\theta,j}) \psi_\theta = i L^S_{\theta,j} \psi_\theta$ belongs to the real linear span of $\{L^S_{\theta,j'} \psi_\theta\}_{j'=1}^n$ because of (5) and the nonsingularity of the matrix $a^k_j(\theta)$. This implies that $\mathcal{P}$ is locally coherent at every point $\theta$.

It is clear from the definition that if $\mathcal{P}$ is locally coherent at $\theta$, then $T_\theta^S(\mathcal{P})$ forms a symplectic space with the symplectic form being the restriction of $[\cdot, \cdot]_{\rho_\theta}$. In particular, the dimensionality of $T_\theta^S(\mathcal{P})$ is necessarily even (say $n = 2m$), and it has a symplectic basis $\{\tilde{L}^S_{\theta,j}\}_{j=1}^{2m}$ satisfying

$$[\tilde{L}^S_{\theta,j}, \tilde{L}^S_{\theta,k}]_{\rho_\theta} = \begin{cases} -1, & \text{if } j \text{ is odd and } k = j + 1 \\ 1, & \text{if } j \text{ is even and } k = j - 1 \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, since $\mathcal{D}_\theta$ is unitary on $T_\theta^S(\mathcal{P})$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\rho_\theta}$, we can take $\{\tilde{L}^S_{\theta,j}\}$ to be orthonormal. Such a basis, which we
shall call a normalized $\rho_\theta$-symplectic basis, satisfies

$$\mathcal{D}_\theta \begin{bmatrix} \hat{L}^S_{\theta,1} \\ \hat{L}^S_{\theta,2} \\ \hat{L}^S_{\theta,3} \\ \vdots \\ \hat{L}^S_{\theta,2m-1} \\ \tilde{L}^S_{\theta,2m} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 1 \\ -1 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \hat{L}^S_{\theta,1} \\ \hat{L}^S_{\theta,2} \\ \hat{L}^S_{\theta,3} \\ \vdots \\ \tilde{L}^S_{\theta,2m-1} \\ \tilde{L}^S_{\theta,2m} \end{bmatrix}.$$ (6)

Thus the SLD tangent space of a coherent model is decomposed into two-dimensional $\mathcal{D}_\theta$-invariant subspaces. This suggests the importance of studying two-dimensional coherent models.

Now, let us characterize a two-dimensional coherent model.

**Theorem 3:** For a two-dimensional pure state model $\mathcal{P} = \{\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta| ; \theta \in \Theta\}$, the following three conditions are equivalent.

1. (α) $\mathcal{P}$ is locally coherent at $\theta$.
2. (β) $L^S_{\theta,1}\psi_\theta$ and $L^S_{\theta,2}\psi_\theta$ are linearly dependent.
3. (γ) $L^A_{\theta,1}\psi_\theta$ and $L^A_{\theta,2}\psi_\theta$ are linearly dependent.

Before going to the proof, we should remark that the condition (β) does not conflict with the fact that $L^S_{\theta,1}$ and $L^S_{\theta,2}$ are linearly independent due to the nondegeneracy of the parametrization $\theta \mapsto \rho_\theta$. Indeed, the linear independence of $\{L^S_{\theta,1}, L^S_{\theta,2}\}$ is concerned with the real linear structure of $L^R_{\sigma}(\rho_\theta)$ and is equivalent to the real linear independence of $\{L^S_{\theta,1}\psi_\theta, L^S_{\theta,2}\psi_\theta\}$. On the other hand, the condition (β) asserts the complex linear dependence of the same vectors.

**Proof:** The proof relies essentially on (4). We only need to show that (α)$\iff$(β), since (β)$\iff$(γ) is obvious from the identity $L^S_{\theta,j}\psi_\theta = -(\mathcal{D}_\theta L^A_{\theta,j})\psi_\theta = -iL^A_{\theta,j}\psi_\theta$. Let $\varphi_j := L^S_{\theta,j}\psi_\theta$, and assume (α) first. Then there exist real numbers $x,y$ such that $\mathcal{D}_\theta L^S_{\theta,1} = xL^S_{\theta,1} + yL^S_{\theta,2}$. This is equivalent to $i\varphi_1 = x\varphi_1 + y\varphi_2$ and leads to (β). Assume (β) in turn. Recalling the real linear independence of $\{\varphi_1, \varphi_2\}$, we see that there exist real numbers $x,y$ satisfying $\varphi_2 = (x + iy)\varphi_1$ with $y \neq 0$. It then follows that $i\varphi_1 = -(x/y)\varphi_1 + (1/y)\varphi_2$ and $\mathcal{D}_\theta L^S_{\theta,1} = -(x/y)L^S_{\theta,1} + (1/y)L^S_{\theta,2}$. Similarly $\mathcal{D}_\theta L^S_{\theta,2}$ is shown to be a real linear combination of $\{L^S_{\theta,1}, L^S_{\theta,2}\}$ and thus (α) is verified.

The following corollary, whose proof is now straightforward from Theorem 3 and (4), offers a mostly useful method to treat group covariant
coherent models as exemplified in Section 6. Moreover the equation (7) in
the corollary reveals a close connection with the conventional definition of
coherent states [9]. Indeed, this fact gave a motive for the nomenclature of
the coherent model.

**Corollary 4:** Let $\mathcal{P} = \{\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|; \theta \in \Theta\}$ be a two-dimensional pure
state model and let $T^A_\theta(\mathcal{P})$ be the real linear span of ALDs $\{L^A_{\theta,1}, L^A_{\theta,2}\}$ at
$\theta$. Then $\mathcal{P}$ is locally coherent at $\theta$ iff there exist nonzero elements $X_1, X_2$ in $T^A_\theta(\mathcal{P})$ satisfying

$$ (X_1 + iX_2)\psi_\theta = 0. \tag{7} $$

Moreover, (7) is also necessary and sufficient for $\{cX_j\}_{j=1,2}$ to form a
normalized $\rho_0$-symplectic basis of $T^S_\theta(\mathcal{P}) = = T^A_\theta(\mathcal{P})$ with a common nor-
malizing constant $c$. Under the condition (7), the linear relations

$$ L^A_{\theta,1} = c_{11}X_1 + c_{12}X_2, \quad L^A_{\theta,2} = c_{21}X_1 + c_{22}X_2 $$

imply

$$ L^S_{\theta,1} = c_{12}X_1 - c_{11}X_2, \quad L^S_{\theta,2} = c_{22}X_1 - c_{21}X_2. $$

4. Generalized RLD Bound

Throughout this section we consider an $n$-dimensional model $\mathcal{P} = \{\rho_\theta\}$ of
general (i.e., not necessarily pure) states satisfying the regularity conditions
(a) and (b) presented in Section 3.

Let $L^2_+(\rho)$ denote the completion of $B(\mathcal{H})$ with respect to the pre-inner
product $(\cdot, \cdot)_\rho$. Since $(X, X)_\rho \leq 2\langle X, X \rangle_\rho$, then $L^2(\rho) \subset L^2_+(\rho)$. The right
logarithmic derivative (RLD) $L^R_{\theta,j}$ in the $j$th direction of a model $\mathcal{P} = \{\rho_\theta\}$,
when it exists, is defined by the requirement that

$$ \frac{\partial}{\partial \theta^j} \text{Tr} \rho_\theta X = (L^R_{\theta,j}, X)_{\rho_\theta}, \quad L^R_{\theta,j} \in L^2_+(\rho_\theta) $$

for all $X \in L^2_+(\rho_\theta)$, or formally $\partial \rho_\theta/\partial \theta^j = (L^R_{\theta,j})^*\rho_\theta = \rho_\theta L^R_{\theta,j}$. The covariance matrix of an arbitrary locally unbiased estimator $M$ is then bounded
from below as

$$ V_\theta[M] \geq (J^R_\theta)^{-1}, \tag{8} $$

where $J^R_\theta = [(L^R_{\theta,j}, L^R_{\theta,k})_{\rho_\theta}]$ is the RLD Fisher information matrix (p. 279
[8]). When a real positive definite matrix $G$ is specified as the weight for
the estimation accuracy, the total deviation is bounded from below as

$$ \text{tr} GV_\theta[M] \geq C^R := \text{tr} G \text{Re} (J^R_\theta)^{-1} + \text{tr} \text{abs} G \text{Im} (J^R_\theta)^{-1}, \tag{9} $$
where $\text{tr abs } A$ denotes the absolute sum of the eigenvalues of matrix $A$, see p. 284 [8]. The RLD thus gives a CR bound and plays a crucial role in optical communication theory [16, 8].

The RLD exists iff there is a constant $c$ such that

$$\left| \frac{\partial}{\partial \theta_j} \text{Tr } \rho_\theta X \right|^2 \leq c(X, X)_{\rho_\theta} \tag{10}$$

for all $X \in \mathcal{B}(\mathcal{H})$. Thus the RLD does not always exist for a model satisfying the weaker condition (b). In particular it never exists for a pure state model $\rho_\theta = |\psi_\theta\rangle \langle \psi_\theta|$. To see this, let us fix a $\theta$ arbitrarily and take a vector $x \in \mathcal{H}$ such that $\langle \psi_\theta| x \rangle = 0$ and $(\partial/\partial \theta_j)|\psi_\theta\rangle \neq 0$. Then $X = |x\rangle \langle \psi_\theta|$ satisfies $(X, X)_{\rho_\theta} = 0$ and $(\partial/\partial \theta_j)\text{Tr } \rho_\theta X \neq 0$. It is, however, important to notice that what is needed in estimation theory is not the RLD itself but the inverse of the RLD Fisher information matrix, as indicated by (8) and (9).

In his book (p. 280 [8]), Holevo has shown that when a model satisfying the regularity conditions (a) and (10) has a $D_\theta$-invariant SLD tangent space, the $(J_\theta^R)^{-1}$ is expressed only in terms of SLDs; so is the CR bound (9). We generalize this result to a wider class of models that satisfy only the weaker conditions (a) and (b).

**Theorem 5:** Suppose we are given an $n$-dimensional model $P = \{\rho_\theta\}$ having a $D_\theta$-invariant SLD tangent space $T_\theta^S(P)$. Then for all locally unbiased measurements $M$ at $\theta$,

$$V_\theta[M] \geq (J_\theta^S)^{-1} + i (J_\theta^S)^{-1} D_\theta (J_\theta^S)^{-1},$$

where $D_\theta = [L_{\theta,j}^S, L_{\theta,k}^S]_{\rho_\theta}$.

**Proof:** Let us introduce a family of inner products on $\mathcal{L}^2(\rho_\theta)$ having a parameter $\varepsilon \in (0, 1]$:

$$(X, Y)_{\rho_\theta}^{(\varepsilon)} = (1 - \varepsilon)(X, Y)_{\rho_\theta} + \varepsilon(X, Y)_{\rho_\theta}.$$ 

Since

$$\varepsilon(X, X)_{\rho_\theta} \leq (X, X)_{\rho_\theta}^{(\varepsilon)} \leq (2 - \varepsilon)(X, X)_{\rho_\theta},$$

there exists, for each $\varepsilon$, a unique operator $L_{\theta,j}^{(\varepsilon)} \in \mathcal{L}^2(\rho_\theta)$ which satisfies

$$\frac{\partial}{\partial \theta_j} \text{Tr } \rho_\theta X = (L_{\theta,j}^{(\varepsilon)}, X)_{\rho_\theta}^{(\varepsilon)}.$$
for all $X \in \mathcal{L}^2(\rho_\theta)$. Then in a quite similar way to the derivation of the quantum Cramér-Rao inequality, we have

$$V_\theta[M] \geq \left( J_\theta^{(\varepsilon)} \right)^{-1}, \quad J_\theta^{(\varepsilon)} = \left[ \langle L_{\theta,j}^{(\varepsilon)}, L_{\theta,k}^{(\varepsilon)} \rangle_{\rho_\theta} \right]. \quad (11)$$

Now observing the identity $(X,Y)^{\rho_\theta} = \langle X,Y \rangle_{\rho_\theta} + i(1-\varepsilon)[X,Y]_{\rho_\theta}$, and using the definition of $\mathcal{D}_{\rho_\theta} = \mathcal{D}_{\theta}$, we see that for all $Y \in \mathcal{L}^2(\rho_\theta)$,

$$\frac{\partial}{\partial \theta} \text{Tr} \rho_\theta Y = \langle L_{\theta,j}^{S}, Y \rangle_{\rho_\theta} = \langle L_{\theta,j}^{(\varepsilon)}, Y \rangle_{\rho_\theta} = \langle \{I + i(1-\varepsilon)\mathcal{D}_\theta \} L_{\theta,j}^{(\varepsilon)}, Y \rangle_{\rho_\theta}.$$

Then $L_{\theta,j}^{S} = \{I + i(1-\varepsilon)\mathcal{D}_\theta \} L_{\theta,j}^{\varepsilon}$, hence $(L_{\theta,j}^{(\varepsilon)}, L_{\theta,j,k}^{(\varepsilon)})_{\rho_\theta} = \langle L_{\theta,j}^{S}, \{I + i(1-\varepsilon)\mathcal{D}_\theta \}^{-1} L_{\theta,j}^{S} \rangle_{\rho_\theta}$. Let us introduce Dirac’s notation $|L_{\theta,j}^{S} \rangle$ for the Hilbert space $(\mathcal{L}^2(\rho_\theta), \langle \cdot, \cdot \rangle_{\rho_\theta})$, and let $\Gamma_\theta := \left[ |L_{\theta,1}^{S} \rangle, \cdots, |L_{\theta,n}^{S} \rangle \right]$. Then $\Gamma_\theta^* \Gamma_\theta = J_\theta^{(\varepsilon)}$ and $\Gamma_\theta^* D_\theta \Gamma_\theta = D_\theta$. And the matrix $J_\theta^{(\varepsilon)}$ can be written in the form $J_\theta^{(\varepsilon)} = \Gamma_\theta^* \{I + i(1-\varepsilon)\mathcal{D}_\theta \}^{-1} \Gamma_\theta$. Thus from the assumption that $\mathcal{T}_\theta^S(\mathcal{P})$ is $\mathcal{D}_\theta$-invariant, the inverse of $J_\theta^{(\varepsilon)}$ is explicitly given by

$$\left( J_\theta^{(\varepsilon)} \right)^{-1} = \left( J_\theta^{S} \right)^{-1} \Gamma_\theta^* \{I + i(1-\varepsilon)\mathcal{D}_\theta \}^{-1} \Gamma_\theta \left( J_\theta^{S} \right)^{-1} = \left( J_\theta^{S} \right)^{-1} + i(1-\varepsilon) \left( J_\theta^{S} \right)^{-1} D_\theta \left( J_\theta^{S} \right)^{-1}. \quad (12)$$

Combining (11) and (12), and taking the limit $\varepsilon \downarrow 0$, we have the theorem.

Theorem 5 asserts that even for a model that does not have the RLDs, the $\lim_{\varepsilon \downarrow 0} (J_\theta^{(\varepsilon)})^{-1}$ indeed gives a generalization of $(J_\theta^R)^{-1}$ as long as the SLD tangent space is $\mathcal{D}_\theta$-invariant. Then by using Theorem 5 and an analogous argument to the derivation of (9), we obtain the CR bound

$$C^R = \text{tr} G \left( J_\theta^{S} \right)^{-1} + \text{tr abs} G \left( J_\theta^{S} \right)^{-1} D_\theta \left( J_\theta^{S} \right)^{-1}, \quad (13)$$

for models each having a $\mathcal{D}_\theta$-invariant SLD tangent space $\mathcal{T}_\theta^S(\mathcal{P})$. This may be called a generalized RLD bound. We will show in the next section that this bound is achievable in a coherent model.

5. Optimal Estimation for Two-Dimensional Coherent Models

We now proceed to a parameter estimation for a pure coherent model. In particular, taking into account the symplectic structure (6) of the SLD tangent space, we restrict ourselves to a two-dimensional case. We note that as long as we are concerned with the achievable CR bound at each point on the model $\{\rho_\theta\}$, we can take the weight as $G = I$ without loss of generality. In
fact, let $M$ be a locally unbiased measurement for the parameter $\theta = (\theta^1, \theta^2)$ and let $p(\hat{\theta}^1, \hat{\theta}^2)d\hat{\theta} = \text{Tr} \rho_\theta M(d\hat{\theta})$ be the corresponding joint distribution. The coordinate transformation $\eta^i = \sum_j h^i_j \theta^j$, where $H = [h^i_j]$ is a real regular matrix, then induces another measurement $N(d\hat{\eta})$ which corresponds to the joint distribution $q(\hat{\eta}^1, \hat{\eta}^2)d\hat{\eta} = p(\hat{\theta}^1, \hat{\theta}^2)d\hat{\theta}$ and is locally unbiased for the parameter $\eta = (\eta^1, \eta^2)$. In this case, $\text{tr} V_\eta[N] = \text{tr} (t^HH)V_\theta[M]$. Thus the parameter estimation for $\theta$ with the weight $G = t^HH$ is equivalent to that for $\eta$ with the weight $I$.

Now suppose we are given a two-dimensional coherent model $\mathcal{P} = \{\rho_\theta; \theta = (\theta^1, \theta^2) \in \Theta\}$. Let \{L^i\} be the dual basis of the SLDs: $L^i = \sum_j J_{ij}^S L^{S\theta,j}$ with $J_{ij}$ being the $(i,j)$ entry of $(J^{S\theta})^{-1}$. Then

$$(J^{S\theta})^{-1} = \begin{bmatrix} \langle L^1, L^1 \rangle_{\rho_\theta} & \langle L^1, L^2 \rangle_{\rho_\theta} \\ \langle L^2, L^1 \rangle_{\rho_\theta} & \langle L^2, L^2 \rangle_{\rho_\theta} \end{bmatrix}$$

and

$$(J^{S\theta})^{-1} D_\theta (J^{S\theta})^{-1} = \begin{bmatrix} 0 & [L^1, L^2]_{\rho_\theta} \\ [L^2, L^1]_{\rho_\theta} & 0 \end{bmatrix}$$

Thus the generalized RLD bound (13) for $G = I$ can be rewritten in the form

$$C^R = \langle L^1, L^1 \rangle_{\rho_\theta} + \langle L^2, L^2 \rangle_{\rho_\theta} + 2 \left| [L^1, L^2]_{\rho_\theta} \right|. \quad (14)$$

We will show that the bound $C^R$ is achievable. In what follows, we fix a $\theta = (\theta^1, \theta^2)$ arbitrarily.

Let us consider a random measurement as follows. We first introduce a linear transformation $\phi: T^{S\theta}(\mathcal{P}) \rightarrow T^{S\theta}(\mathcal{P})$ by

$$\phi(X) = \langle L^1, X \rangle_{\rho_\theta} L^1 + \langle L^2, X \rangle_{\rho_\theta} L^2.$$

Since $\phi$ is symmetric and positive definite, it has positive eigenvalues $\lambda_1, \lambda_2$, and mutually orthogonal unit eigenvectors $A_1, A_2$ satisfying $\phi(A_\nu) = \lambda_\nu A_\nu, \quad \nu = 1, 2$. We next take positive numbers $p_1, p_2$ satisfying $p_1 + p_2 = 1$. Now letting

$$\int \xi E_\nu(d\xi), \quad \nu = 1, 2$$

be the spectral decompositions of arbitrarily fixed representatives of $A_\nu$, we define a generalized measurement

$$M(\nu, d\xi) = p_\nu E_\nu(d\xi).$$
This has the following physical interpretation: Select one of the two “observables” $A_1, A_2$ according to the probability $p_1, p_2$, respectively, and measure it in a usual sense.

Now suppose we have selected $A_\nu$ and have obtained an outcome $\xi$. We identify this result with a pair of real quantities

$$\hat{\theta}^i(\nu, \xi) = \theta^i + \frac{\xi}{p_\nu} \langle L^i, A_\nu \rangle_{\rho_\theta}, \quad i = 1, 2.$$  

The pair $\{\hat{\theta}^i(\nu, \xi)\}_{i=1,2}$ satisfies the local unbiasedness condition at $\theta$:

$$\sum_{\nu=1}^2 \int \hat{\theta}^i(\nu, \xi) \, \text{Tr} \rho_\theta M(\nu, d\xi) = \theta^i, \quad i = 1, 2 \tag{15}$$

$$\sum_{\nu=1}^2 \int \hat{\theta}^i(\nu, \xi) \, \frac{\partial}{\partial \theta^j} \text{Tr} \rho_\theta M(\nu, d\xi) = \delta^i_j, \quad i, j = 1, 2. \tag{16}$$

To prove (15), we used the fact that $A_\nu \in T^S_\theta(\mathcal{P})$, i.e., $\langle I, A_\nu \rangle_{\rho_\theta} = 0$. To prove (16), observe that

$$\int \xi \frac{\partial}{\partial \theta^j} \text{Tr} \rho_\theta E^\nu(\xi) \, d\xi = \langle L^S_{\theta,j}, A_\nu \rangle_{\rho_\theta},$$

so that the left-hand side of (16) becomes

$$\sum_{\nu=1}^2 \langle L^i, A_\nu \rangle_{\rho_\theta} \langle L^S_{\theta,j}, A_\nu \rangle_{\rho_\theta} = \langle L^i, L^S_{\theta,j} \rangle_{\rho_\theta} = \delta^i_j.$$  

With this measurement $M$,

$$\text{tr} V_\theta[M] = \sum_{\nu=1}^2 \int \left[ \left( \hat{\theta}^1(\nu, \xi) - \theta^1 \right)^2 + \left( \hat{\theta}^2(\nu, \xi) - \theta^2 \right)^2 \right] \text{Tr} \rho_\theta M(\nu, d\xi)$$

$$= \sum_{\nu=1}^2 \frac{1}{p_\nu} \left[ \langle L^1, A_\nu \rangle_{\rho_\theta}^2 + \langle L^2, A_\nu \rangle_{\rho_\theta}^2 \right].$$

In the second equality, we used the fact that

$$\int \xi^2 \, \text{Tr} \rho_\theta E^\nu(\xi) \, d\xi = \langle A_\nu, A_\nu \rangle_{\rho_\theta} = 1.$$  

Since, for given $\mu_1, \mu_2 > 0$, $\mu_1/p_1 + \mu_2/p_2$ takes the minimum $(\sqrt{\mu_1} + \sqrt{\mu_2})^2$
at $p_\nu = \frac{\sqrt{\mu_\nu}}{\sqrt{\mu_1} + \sqrt{\mu_2}}$, we see

$$
\min_{\{p_\nu\}} \text{tr} V_0[M] = \left[ \sqrt{\langle L^1, A_1 \rangle_{\rho_\theta}^2 + \langle L^2, A_1 \rangle_{\rho_\theta}^2 + \sqrt{\langle L^1, A_2 \rangle_{\rho_\theta}^2 + \langle L^2, A_2 \rangle_{\rho_\theta}^2}} \right]^2 \\
= \left[ \sqrt{\langle A_1, \phi(A_1) \rangle_{\rho_\theta}} + \sqrt{\langle A_2, \phi(A_2) \rangle_{\rho_\theta}} \right]^2 \\
= \left[ \sqrt{\lambda_1} + \sqrt{\lambda_2} \right]^2 \\
= \langle L^1, L^1 \rangle_{\rho_\theta} + \langle L^2, L^2 \rangle_{\rho_\theta} + 2 \sqrt{\langle L^1, L^1 \rangle_{\rho_\theta} \langle L^2, L^2 \rangle_{\rho_\theta} - \langle L^1, L^2 \rangle_{\rho_\theta}^2}. (17)
$$

The last equality follows from the fact that the trace $\lambda_1 + \lambda_2$ and the determinant $\lambda_1 \lambda_2$ of the linear transformation $\phi$ are independent of the choice of the basis which represents $\phi$ in a matrix form.

The random measurement presented above was first introduced in [13] by one of the present authors. In that paper, it was also shown that the problem of finding the achievable CR bound for an arbitrary two-parameter faithful spin 1/2 model can be reduced to an easy minimization problem. Interestingly, the explicit solution of the minimization problem, i.e., the achievable CR bound, turns out to be identical to the quantity (17), although the model treated there is not pure nor has in general a $D_\rho$-invariant tangent space.

Now we establish the relation between (14) and (17) for a coherent model.

**Theorem 6:** For a two-dimensional coherent model $\{\rho_\theta = |\psi_\theta \rangle \langle \psi_\theta|\}$, the lower bound (14) is identical to (17). In other words, the generalized RLD bound (14) is achievable.

**Proof:** By Theorem 3, $L^1 \psi_\theta$ and $L^2 \psi_\theta$ are linearly dependent. Therefore

$$
\det \left[ \begin{array}{cc}
\langle L^1 \psi_\theta | L^1 \psi_\theta \rangle & \langle L^1 \psi_\theta | L^2 \psi_\theta \rangle \\
\langle L^2 \psi_\theta | L^1 \psi_\theta \rangle & \langle L^2 \psi_\theta | L^2 \psi_\theta \rangle
\end{array} \right] = 0,
$$

which leads to

$$(\text{Im} \langle L^1 \psi_\theta | L^2 \psi_\theta \rangle)^2 = \langle L^1 \psi_\theta | L^1 \psi_\theta \rangle \langle L^2 \psi_\theta | L^2 \psi_\theta \rangle - \langle L^1 \psi_\theta | L^2 \psi_\theta \rangle^2.$$

By (1) and (2), this can be read as

$$
||[L^1, L^2]_{\rho_\theta}||^2 = \langle L^1, L^1 \rangle_{\rho_\theta} \langle L^2, L^2 \rangle_{\rho_\theta} - \langle L^1, L^2 \rangle_{\rho_\theta}^2,
$$

which proves the theorem. ■
It should be noted that a more convincing result has been obtained by Matsumoto [11]. He proved that the CR bound (13) is achievable for a $2m$-dimensional coherent model with an arbitrary weight $G$.

It is also worth noting that the achievability of (14) is closely related to the Heisenberg uncertainty relation. By a coordinate transformation, we can assume that the SLD Fisher information matrix is diagonal at a fixed $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$. Then there exist nonzero real numbers $c_1, c_2$ and normalized $\rho_\theta$-symplectic basis $\{\tilde{L}_1^S, \tilde{L}_2^S\}$ such that $L_j^S = c_j \tilde{L}_j^S$. Then $L_j = \tilde{L}_j^S / c_j$, and by (7)

$$(c_1 L^1 + i c_2 L^2) \psi_\theta = 0.$$ 

This is nothing but the equality condition for the Heisenberg uncertainty relation. So we have

$$\langle L^1, L^1 \rangle_{\rho_\theta} \langle L^2, L^2 \rangle_{\rho_\theta} = |[L^1, L^2]_{\rho_\theta}|^2.$$ 

This equation, combined with the assumption that $\langle L^1, L^2 \rangle_{\rho_\theta} = 0$, gives another proof of Theorem 6 for an orthogonal parametrization at $\rho_\theta$.

6. Examples

In this section we calculate the achievable CR bounds for canonical and spin coherent models. Throughout this section, adjoint operators and complex conjugate numbers are denoted by $^\dagger$ and $^*$, respectively, according to the convention in physics. Also we use the same letter for both a square summable operator and the corresponding element in $\mathcal{L}_h^2(\rho)$.

6.1. Canonical squeezed state model

The canonical squeezed state [2, 15] is defined by

$$\rho_{q,p} = D(q,p)|0\rangle\langle0| D^\dagger(q,p), \quad (q,p \in \mathbb{R}),$$

where $D(q,p) = \exp(za^\dagger - z^*a)$ denotes the shift operator with $z = (q + ip)/\sqrt{2}$, and $a$ and $a^\dagger$ are annihilation and creation operators, respectively, with $a = (Q + iP)/\sqrt{2}$. Further $|0\rangle_{\xi} = \exp[(\xi a^\dagger^2 - \xi^*a^2)/2]|0\rangle$ is the squeezed vacuum with $|0\rangle$ the Fock vacuum, and $\xi$ a complex number which represents the squeezing ratio: let $\xi = se^{i\theta}$.

Comparing the identity $b|z\rangle_{\xi} = \beta |z\rangle_{\xi}$ with Corollary 4, where $|z\rangle_{\xi} = D(q,p)|0\rangle_{\xi}$, $b = a \cosh s - a^\dagger e^{i\theta} \sinh s$, and $\beta = z \cosh s - z^*e^{i\theta} \sinh s$, we
see that $\rho_{q,p}$ is a two-dimensional coherent model, and a normalized $\rho_{q,p}$-symplectic basis is given by

\[
\begin{align*}
\tilde{L}_1^S &= \sqrt{2}[(Q-qI)(\cosh s - \cos \theta \sinh s) - (P-pI) \sin \theta \sinh s], \\
\tilde{L}_2^S &= \sqrt{2}[(P-pI)(\cosh s + \cos \theta \sinh s) - (Q-qI) \sin \theta \sinh s].
\end{align*}
\]

The SLDs at $\rho_{q,p}$ are calculated by operating $-D_{q,p}$ to ALDs at $\rho_{q,p}$. By expanding ALDs $L_q^A = 2(P-pI), L_p^A = -2(Q-qI)$ into linear combinations of $\tilde{L}_1^S, \tilde{L}_2^S$, and using the relations $D_{q,p}\tilde{L}_1^S = \tilde{L}_2^S, D_{q,p}\tilde{L}_2^S = -\tilde{L}_1^S$, we have

\[
\begin{align*}
L_q^S &= 2[(Q-qI)(\cosh 2s - \cos \theta \sinh 2s) - (P-pI) \sin \theta \sinh 2s], \\
L_p^S &= 2[(P-pI)(\cosh 2s + \cos \theta \sinh 2s) - (Q-qI) \sin \theta \sinh 2s].
\end{align*}
\]

The corresponding SLD Fisher information matrix becomes

\[
J_{q,p}^S = 2 \begin{bmatrix}
\cosh 2s - \cos \theta \sinh 2s & -\sin \theta \sinh 2s \\
-\sin \theta \sinh 2s & \cosh 2s + \cos \theta \sinh 2s
\end{bmatrix}.
\]

Then from (17), we have

\[
\min_M \text{tr} V_{q,p}[M] = \cosh 2s + 1.
\]

### 6.2. Spin coherent state model

The spin coherent state $[1,14]$ in the spin $j$ representation is defined by

\[
\rho_{\theta,\varphi} = R(\theta,\varphi)|j\rangle\langle j|R^\dagger(\theta,\varphi), \quad (0 \leq \theta \leq \pi, 0 \leq \varphi < 2\pi),
\]

where $(\theta, \varphi)$ is the polar coordinate system (the north pole is $\theta = 0$ and $x$-axis corresponds to $\varphi = 0$), $R(\theta, \varphi) = \exp[i\theta(J_x \sin \varphi - J_y \cos \varphi)]$ the rotation through an angle $-\theta$ about an axis $(\sin \varphi, -\cos \varphi, 0)$, and $|j\rangle$ the highest weight state with respect to $J_z$ that corresponds to the north pole.

Since $J_+|j\rangle = (J_x + iJ_y)|j\rangle = 0$, we find that $\rho_{\theta,\varphi}$ is a two-dimensional coherent model, and a normalized $\rho_{0,0}$-symplectic basis is $\tilde{L}_1^S(0,0) = \sqrt{2/j}J_x$, $\tilde{L}_2^S(0,0) = \sqrt{2/j}J_y$. A normalized $\rho_{\theta,\varphi}$-symplectic basis is then calculated as

\[
\tilde{L}_k^S(\theta, \varphi) = R(\theta, \varphi)\tilde{L}_k^S(0,0)R^\dagger(\theta, \varphi),
\]

where $k = 1, 2$.

On the other hand, the generators of rotations about axes $(\sin \varphi, -\cos \varphi, 0)$ and $(\cos \varphi, \sin \varphi, 0)$ at $\theta = 0$ are $i(J_x \sin \varphi - J_y \cos \varphi)$
and $i(J_x \cos \varphi + J_y \sin \varphi)$, respectively. Therefore ALDs for the model at $\rho_{\theta,\varphi}$ are given by

$$L^A_{\theta}(\theta, \varphi) = R(\theta, \varphi) \{ -2(J_x \sin \varphi - J_y \cos \varphi) \} R^\dagger(\theta, \varphi)$$

$$= -\sqrt{2}j \left\{ \tilde{L}^S_1(\theta, \varphi) \sin \varphi - \tilde{L}^S_2(\theta, \varphi) \cos \varphi \right\},$$

$$L^A_{\varphi}(\theta, \varphi) = R(\theta, \varphi) \{ -2(J_x \cos \varphi + J_y \sin \varphi) \sin \theta \} R^\dagger(\theta, \varphi)$$

$$= -\sqrt{2}j \left\{ \tilde{L}^S_1(\theta, \varphi) \sin \theta \cos \varphi + \tilde{L}^S_2(\theta, \varphi) \sin \theta \sin \varphi \right\}.$$ 

The SLDs at $\rho_{\theta,\varphi}$ are calculated by operating $-D_{\theta,\varphi}$ to ALDs, to obtain

$$L^S_{\theta}(\theta, \varphi) = \sqrt{2}j \left\{ \tilde{L}^S_1(\theta, \varphi) \cos \varphi + \tilde{L}^S_2(\theta, \varphi) \sin \varphi \right\},$$

$$L^S_{\varphi}(\theta, \varphi) = -\sqrt{2}j \left\{ \tilde{L}^S_1(\theta, \varphi) \sin \theta \sin \varphi - \tilde{L}^S_2(\theta, \varphi) \sin \theta \cos \varphi \right\}.$$ 

Since $\rho_{\theta,\varphi}$-symplectic basis $\{ \tilde{L}^S_k(\theta, \varphi) \}_{k=1,2}$ is orthonormal, the SLD Fisher information matrix and the matrix $D$ are easily calculated:

$$J^S_{\theta,\varphi} = \begin{bmatrix} 2j & 0 \\ 0 & 2j \sin^2 \theta \end{bmatrix}, \quad D_{\theta,\varphi} = \begin{bmatrix} 0 & -2j \sin \theta \\ 2j \sin \theta & 0 \end{bmatrix}.$$ 

We thus have

$$\min_M \text{tr} V_{\theta,\varphi}[M] = \frac{1}{2} \left( 1 + \frac{1}{\sin \theta} \right)^2.$$ 

7. Conclusions

We introduced a class of quantum pure state models called the coherent models. They are characterized by a symplectic structure of the tangent space, and have a close connection with the conventional generalized coherent states in mathematical physics. A Cramér-Rao type bound for a coherent model was derived by an analogous argument to the derivation of the right logarithmic derivative bound. Moreover, by an argument of random measurement, this lower bound was found to be achievable.

Acknowledgment

We thank Keiji Matsumoto for helpful suggestions, with which the manuscript has been improved as compared with the early version [3] of this paper.
References

CHAPTER 20
A Geometrical Approach to Quantum Estimation Theory
Keiji Matsumoto

1. Introduction

1.1. The purposes of the thesis

The most important purpose of the thesis is pursuit for the geometrical theory of statistical estimation of the quantum mechanical state.

In the statistical theory of the probability distribution, S. Amari and his coworkers have formulated a geometrical theory, so called information geometry, and successfully applied to various statistical problems [3]. H. Nagaoka, one of Amari’s coworkers, pointed out that the duality between e- and m-connections sits at the heart of the information geometry. He also formulated quantum information geometry by use of this idea of the mutually dual connections, and applied to characterization of the model which has the efficient estimator [16, 17].

However, his information geometry does not give any insight into the problem of determination of the attainable CR (Cramér-Rao) type bound, nor characterize the condition for the SLD CR bound, which is of special interest for some reasons, is attainable. After all, Nagaoka’s geometry deal with the global properties of the model, while the attainable CR type bound is related local properties of the model. Hence, it seems that another geometric structure is needed for the thorough description of the quantum estimation theory.

On the other hand, Berry’s phase, discovered by M. V. Berry as the non-integrable phase factor in the adiabatic motion [4], is naturally understood as a curvature of the natural connection in the principle fiber bundle over

This chapter is reprinted from K. Matsumoto’s Doctoral thesis “A Geometrical Approach to Quantum Estimation Theory”. This doctoral thesis contains the topics, Estimation of the temperature, Nagaoka’s information geometry, Uncertainty principle in view of quantum estimation theory, Time-energy uncertainty in view of hypothesis test, w-connection and e-connection, Attainability of SLD CR bound. But, these parts are omitted in this reprint with author’s permission.
the space of pure states whose structure group is $U(1)$ [1]. In 1986, Uhlmann generalized the geometry to the space of mixed states. Though the physical meaning of Uhlmann’s geometry is not known, Berry’s phase is applied to the explanation of various phenomena [22].

The author conjectures that Berry-Uhlmann’s curvature reflects local properties of the model. To prove the statement, it is needed to determine the attainable CR type bound for arbitrary models, which is far out of our reach. However, for the 2-parameter pure state model, the author presents complete answer to the problem. In addition, for the faithful state model and pure state model, it is shown that SLD CR bound is attainable if and only if the model is free of Berry-Uhlmann curvature.

Furthermore, we try a kind of unification of the two geometries, Nagaoka’s information geometry and Uhlmann’s parallelism.

Second most important purpose is the determination of the attainable CR type bound, and the development of new methodology for that purpose. In the pure state models, this purpose is achieved, though not completely, to large extent. A new methodology direct approach is formulated and successfully applied to the 2-parameter pure state model, the coherent model. For the arbitrary pure state model, calculated is the attainable CR type bound whose weight matrix is the SLD Fisher information matrix. Looking back, no one has ever determined the attainable CR type bound for this wide range of models.

Third, some considerations about such physical problems as the uncertainty principle are done. The time-energy uncertainty is nicely formulated as a hypothesis test, and the position-momentum uncertainty as a estimation of the mean values of the position and the momentum operators. In this formulation of the position-momentum uncertainty, it is shown that the mean values of the position and the momentum operators are simultaneously estimated up to arbitrarily high efficiency, if the particle is prepared carefully.

1.2. Organization of the thesis

The thesis is divided into three parts: the faithful model theory, the pure state theory, the general model theory; The reason for this organization is that the extent of the achievement of the purposes is different in these three cases.

Before these three parts, section 2 gives brief review of the estimation theory of probability distributions, the quantum mechanical theory of the
measurement, and the quantum estimation theory, and section 3 gives the geometrical and the estimation-theoretical framework, commonly used in any of the following three parts.

2. Preliminaries

In this section, statistical estimation theory and quantum mechanics are reviewed briefly. For the thorough description of estimation theory, see, for example, Ref. [14]. As for quantum mechanics, see Ref. [20], or other textbooks. Some basic concepts in quantum estimation theory are introduced also.

2.1. Classical estimation theory

Throughout this section, the usual estimation theory, or the estimation theory of the probability distribution is called classical estimation theory, in the sense that the theory is not quantum mechanical.

The theme of the classical estimation theory is identification of the probability distribution from which the $N$ data $x_1, x_2, ..., x_N$ is produced. Usually, the probability distribution is assumed to be a member of a model, or a family

$$\mathcal{M} = \{p(x|\theta) | \theta \in \Theta \subset \mathbb{R}^m \}$$

of probability distributions and that the finite dimensional parameter $\theta \in \Theta \subset \mathbb{R}^m$ is to be estimated statistically.

Unbiased estimator $\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_N)$ of parameter $\theta$ is the estimate which satisfies

$$E_\theta[\hat{\theta}] \equiv \int dx_1 dx_2 ... dx_N \hat{\theta}(x_1, x_2, ..., x_N) \prod_{i=1}^{N} p(x_i | \theta) = \theta,$$  \hspace{1cm} (1)

that is, the estimate which gives the true value of parameter in average. For the technical reason, we also define locally unbiased estimator $\hat{\theta}$ at $\theta_0$ by

$$E_{\theta_0}[\hat{\theta}] = \theta_0, \quad \partial_j E_{\theta}[\hat{\theta}^i] \bigg|_{\theta = \theta_0} = \delta^i_j.$$  

The estimator is unbiased iff it is locally unbiased at every $\theta \in \Theta$.

For the variance of locally unbiased estimator at $\theta$, the following theorem gives bound of efficiency of the estimation.
Theorem 1: (Cramér-Rao inequality) For any locally unbiased estimate \( \hat{\theta} \) at \( \theta \),

\[
V_{\theta}[\hat{\theta}] \geq \frac{1}{N} J^{-1}(\theta).
\]

(2)

Here, \( N \) is the number of the data and \( J(\theta) \) is \( m \times m \) real symmetric matrix defined by

\[
J(\theta) \equiv \left[ \int dx p(x|\theta) \partial_i \ln p(x|\theta) \partial_j \ln p(x|\theta) \right],
\]

(3)

where \( \partial_i \) stands for \( \partial/\partial \theta^i \).

The best estimator, or the estimator \( \hat{\theta} \) satisfying (4), is given by

\[
\hat{\theta}^i(x_1, ..., x_N) = \hat{\theta}_0^i(x_1, ..., x_N) \equiv \theta^i + \sum_{j=1}^{m} [J^{-1}(\theta)]^{ij} \partial_j \ln \prod_{k=1}^{N} p(x_k|\theta).
\]

\( J(\theta) \) is called Fisher information matrix, because the larger the \( J(\theta) \) is, the more precise estimate can be done with the same number of data. Metaphorically speaking, we obtain as much information as \( J(\theta) \) per data.

Actually, as easily seen by putting \( N = 1 \) in Cramér-Rao (CR) inequality, we can obtain \( J(\theta) \) as the minimum variance of locally unbiased estimate when only one data is given.

\[
V_{\theta}[\hat{\theta}] \geq J^{-1}(\theta).
\]

(4)

The trouble with the (4) is that the best estimator \( \hat{\theta}_0(\theta) \) is dependent on the true value of the parameter \( \theta \), which is unknown to us. When the true value of parameter is not \( \theta_0 \), the estimate \( \hat{\theta}_0(\theta_0) \) is not even locally unbiased at the true value of the parameter. To avoid this dilemma, we give up with the unbiased estimator, and focus on the consistent estimator defined by

\[
\lim_{N \to \infty} E_{\theta} \left[ \hat{\theta}(x_1, x_2, ..., x_N) \right] = \theta.
\]

For the consistent estimator, we also have the following theorem.

Theorem 2: If the estimator is consistent,

\[
V_{\theta}[\hat{\theta}] \geq \frac{1}{N} J^{-1}(\theta) + o \left( \frac{1}{N} \right)
\]

(5)

holds true.

The maximum likelihood estimator \( \hat{\theta}_{MLE} \), which is defined by,

\[
\hat{\theta}_{MLE} \equiv \arg\max \left\{ \sum_{j=1}^{N} \ln p(x_j|\theta) \mid \theta \in \Theta \subset \mathbb{R}^m \right\},
\]
is consistent and achieves the equality in (5).

Notice that to obtain $\hat{\theta}_{\text{MLE}}$, we need no information about the true value of the parameter beforehand. Hence, the Fisher information matrix is a good measure of the efficiency of the optimal consistent estimator.

2.2. Quantum mechanics and measurement theory

In the quantum mechanics, the state of physical system is described by the density operator $\rho$, which is a nonnegative Hermitian operator whose trace is equal to 1, in a separable Hilbert space $\mathcal{H}$, whose dimension is denoted by $d \leq \infty$ hereafter. We denote by $\mathcal{P}(\mathcal{H})$ the space of density operators in $\mathcal{H}$, by $\mathcal{P}_r(\mathcal{H})$ the space of density operators whose rank is $r$, and by $\mathcal{P}_+(\mathcal{H})$ the space of strictly positive definite density operators. $\mathcal{P}(\mathcal{H})$, $\mathcal{P}_r(\mathcal{H})$, and $\mathcal{P}_+(\mathcal{H})$ are often simply denoted by $\mathcal{P}$, $\mathcal{P}_r$, and $\mathcal{P}_+$, respectively.

Let $\Omega$ be a space of all possible outcomes of an experiment, and $\sigma(\Omega)$ be a $\sigma$-field in $\Omega$. When the density operator of the system is $\rho$, the probability that the data $\omega \in \Omega$ lies in $B \in \sigma(\Omega)$ writes

$$\Pr\{\omega \in B|\rho\} = \text{tr}\rho M(B),$$

(6)

by use of the map $M$ from $\sigma(\Omega)$ to nonnegative Hermitian operator which satisfies

$$M(\phi) = O, M(\Omega) = I, M\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} M(B_i) \left(B_i \cap B_j = \phi, i \neq j\right)$$

(7)

so that (6) define a probability measure (see Ref. [11], p. 53 and Ref. [12], p. 50). We call the map $M$ the measurement, because there always exists a physical experiment corresponding to the map $M$ which satisfies (7) [19, 23].

2.3. Unbiased estimator in quantum estimation theory

The purpose of the quantum estimation is to identify the density operator of the given physical system from the data obtained by the appropriately designed experiment. For simplicity, we usually assume that the density operator is a member of a model, or a manifold of $\mathcal{M} = \{\rho(\theta)|\theta \in \Theta \subset \mathbb{R}^m\} \subset \mathcal{P}$, and that the parameter $\theta$ is to be estimated statistically. For example, $\mathcal{M}$ is the set of spin states with given wave function part and unknown spin part.

To estimate the parameter, we perform an experiment to obtain the data $\omega$ by which we calculate an estimator $\hat{\theta}$ by estimator $\hat{\theta}(\omega)$. A pair
$(\hat{\theta}, M, \Omega)$ of a space $\Omega$ of data, a measurement $M$, and an estimator $\hat{\theta}(\cdot)$ is also called an estimator. The expectation of $f(\omega)$ with respect to the probability measure (6) is denoted by $E_{\theta}[f(\omega)|M]$.

We have seen that the locally unbiased estimator played a key role in classical estimation theory. Hence we try to keep the same track also in the quantum estimation theory.

The estimator $(\hat{\theta}, M, \Omega)$ is said to be unbiased if
\[ E_{\theta}[\hat{\theta}(\omega)|M] = \theta \]
holds for all $\theta \in \Theta$. If (8) and
\[ \partial_i E_{\theta}[\hat{\theta}^j(\omega)|M] = \delta^j_i \quad (i, j = 1, ..., m) \]
hold at a particular $\theta$, $(\hat{\theta}, M, \Omega)$ is called locally unbiased at $\theta$.

It is also reasonable to include calculation of the estimate from data into the process of measurement. In this point of view, the estimate $\hat{\theta}$ itself is produced by the measurement process, and the data space $\Omega$ is $\mathbb{R}^m$. Therefore, by the term ‘estimator’ we also mean the measurement which takes value on $\mathbb{R}^m$. In this case, the unbiased estimator is a measurement which takes value on $\mathbb{R}^m$ which satisfies
\[ E_{\theta}[M] = \theta \]
holds for all $\theta \in \Theta$, where,
\[ E_{\theta}[M] \equiv \int \hat{\theta} \text{tr} \rho(\theta) M(d\hat{\theta}). \]
If (9) and
\[ [\partial_i E_{\theta}[M]]^j = \delta^j_i \quad (i, j = 1, ..., m) \]
hold at a particular $\theta$, $M$ is called locally unbiased at $\theta$. We denote by $V_\theta[M]$ the covariance matrix of the estimator $M$ when the true value of the parameter is $\theta$.

Obviously, these two definition of the estimator are equivalent. Therefore, in some situations, we prefer the former to the latter, while in other situations the latter is preferred for the sake of simplicity.

3. Conceptual Framework

3.1. Horizontal lift and SLD

In this section, except for the pure state model theory, $d \equiv \text{dim } \mathcal{H}$ is assumed to be finite for the sake of clarity. The author believes the essence of the discussion will not be damaged by this restriction.
Let $\mathcal{W}_r$ be the space of $d \times r$ complex and full-rank matrix $W$ such that
\[ \text{tr} WW^* = 1, \]
$\mathcal{P}_r$, the space of density operators whose rank is $r$, and $\pi$ the map from $\mathcal{W}_r$ to $\mathcal{P}_r$ such that
\[ \rho = \pi(W) \equiv WW^*. \]
Because $\pi(WU)$ is identical to $\pi(W)$ iff $U$ is a $r \times r$ unitary matrix, it is natural to see the space $\mathcal{W}_r$ as the total space of the principal fiber bundle with the base space $\mathcal{P}_r$ and the structure group $U(d)$ [13]. One possible physical interpretation of $W$ is a representation of a state vector $|W\rangle$ in a bigger Hilbert space $\mathcal{H} \otimes \mathcal{H}'$. Here, the dimension of $\mathcal{H}'$ is $r$ and the operation $\pi(*)$ corresponds to the partial trace of $|W\rangle\langle W|$ over $\mathcal{H}'$.

In this subsection, basic concepts about the tangent bundle $T(\mathcal{W}_r)$ over $\mathcal{W}_r$, which is a real manifold with the real parameter $\zeta = (\zeta^1, ..., \zeta^{2rd-1})^T$, are introduced.

The matrix representation $M(\partial/\partial \zeta^i)$ of the tangent vector $\partial/\partial \zeta^i$ (throughout the thesis, the tangent vector is understood as the differential operator) is a $d \times r$ complex matrix such that
\[ M\left( \frac{\partial}{\partial \zeta^i} \right) \equiv 2 \frac{\partial}{\partial \zeta^i} W(\zeta). \]
The real span of the matrix representations is
\[ \{ X \mid \text{Re tr} XW^*(\zeta) = 0, X \in M(d, r, \mathbb{C}) \}. \]
We introduce the inner product $\langle \langle *, * \rangle \rangle_W$ to $T(\mathcal{W}_r)$ such that,
\[ \langle \langle \hat{X}, \hat{Y} \rangle \rangle_W \equiv \sum_{i,j} (\text{Re}(M\hat{X})_{ij}\text{Re}(M\hat{Y})_{ij} + \text{Im}(M\hat{X})_{ij}\text{Im}(M\hat{Y})_{ij}) \]
\[ = \text{Re tr} \left( (M\hat{X})(M\hat{Y})^* \right), \]
which is invariant under the action of $U \in U(r)$ to the matrix representation of the tangent vector from right side,
\[ \langle \langle (M\hat{X})U, (M\hat{Y})U \rangle \rangle_W = \langle \langle M\hat{X}, M\hat{Y} \rangle \rangle_W, \quad U \in U(r), \]

Let us decompose $T_\mathcal{W}(\mathcal{W}_r)$ into the direct sum of the horizontal subspace $\mathcal{L}\mathcal{S}_\mathcal{W}$ and the vertical subspace $\mathcal{K}_\mathcal{W}$ where $\mathcal{L}\mathcal{S}_\mathcal{W}$ is defined by
\[ \mathcal{L}\mathcal{S}_\mathcal{W} \equiv \{ \hat{X} \mid W^*(M\hat{X}) = (M\hat{X})^*W \}, \quad (10) \]
and $\mathcal{K}_W$ is the orthogonal complement space $T_W(W_r) \ominus \mathcal{L}S_W$ with respect to the inner product $\langle\langle *, * \rangle \rangle_W$. $\hat{X} \in \mathcal{K}_W$ satisfies

$$\langle (M \hat{X})^*, W(M \hat{X})^* \rangle = 0,$$

or its equivalence,

$$\pi_*(\hat{X}) = 0,$$

where $\pi_*$ is the differential map of $\pi$. A member of the horizontal subspace and the vertical subspace are called a horizontal vector and vertical vector, respectively. The image of $\hat{X} \in T_W(W_r)$ by the projection onto the horizontal subspace $\mathcal{L}S_W$ is called the horizontal component, while the image by the projection onto the vertical subspace $\mathcal{K}_W$ is called vertical component.

The horizontal lift $h_W$ is a mapping from $T_{\pi(W)}(P_r)$ to $T_W(W_r)$ such that

$$\pi_*\left(h_W(X)\right) = X, \quad h_W(X) \in \mathcal{L}S_W.$$

Because of the following theorem, the matrix representation of the horizontal lift $\pi_*(h_W(X))$ is a representation of the tangent vector $X \in T_{\pi(W)}(P_r)$.

**Theorem 3:** $h_W$ is an isomorphism from $T_{\pi(W)}(P_r)$ to $\mathcal{L}S_W$.

**Proof:** First, notice that for any $\hat{Y} \in T_W(W_r)$, $W + \varepsilon M(\hat{Y})$ also is a member of $W_r$, if $\varepsilon$ is small enough. Therefore, we have

$$\pi_*(T_W(W_r)) \subset T_{\pi(W)}(P_r).$$

Second, we prove that the map $\pi_*|_{\mathcal{L}S_W}$ is a one to one map from $\mathcal{L}S_W$ to $T_{\pi(W)}(P_r)$. For that, it is sufficient to prove that $\hat{X} = 0$ when $\hat{X} \in \mathcal{K}_W$. This statement is proved to be true because $\mathcal{K}_W$ is orthogonal to $\mathcal{L}S_W$.

Finally, checking the dimension of $T_{\pi(W)}(P_r)$ is equal to $\mathcal{L}S_W$, we have the theorem.

Using the horizontal lift, the inner product $\langle *, * \rangle$ in $T(P_r)$ is deduced from $\langle\langle *, * \rangle \rangle$:

$$\langle X, Y \rangle_{\pi(W)} = \langle\langle h_W(X), h_W(Y) \rangle \rangle_W.$$

The horizontal lift $h$ satisfies the following equality so that the above definition of the inner product $\langle *, * \rangle$ is self-consistent:

$$\langle\langle h_W X, h_W Y \rangle \rangle_W = \langle\langle h_{UW} X, h_{UW} Y \rangle \rangle_W, \quad (U, U' \in U(n)).$$
The symmetrized logarithmic derivative (SLD, in short) of \( X \in T_{\pi(W)}(P_r) \) is the Hermitian operator \( L^S_X \) in \( \mathcal{H} \) defined by the equation

\[
X \rho(\theta) = \frac{1}{2}(L^S_X \rho(\theta) + \rho(\theta)L^S_X),
\]

where \( \theta \) is a real parameter which is assigned to a member of \( P_r \). If the density operator is strictly positive, SLD is uniquely defined by (13). \( L^S_{\partial/\partial \theta^i} \) is often denoted simply by \( L^S_i \).

SLD is closely related to the horizontal lift by the following equation:

\[
M \left( h_X W \right) = L^S_X W.
\]

### 3.2. Definition of Uhlmann’s parallelism

Berry’s phase, by far confirmed by several experiments, is a holonomy of a natural connection in the line bundle over the space of pure states [1, 4]. In 1986, Uhlmann generalized the theory to include mixed states in the Hilbert space \( \mathcal{H} \) [24, 25, 26]. Throughout this section, for the sake of clarity, \( d \equiv \dim \mathcal{H} \) is assumed to be finite. For notational simplicity, the argument \( \theta \) is omitted, as long as the omission is not misleading.

Define a horizontal lift of a curve \( C = \{\rho(t)|t \in \mathbb{R}\} \) in \( P_r \) as a curve \( C_h = \{W(t)|t \in \mathbb{R}\} \) in \( W_r \) which satisfies \( C = \pi(C_h) \) and

\[
\frac{dW(t)}{dt} = M \left( h_{W(t)} \left( \frac{d}{dt} \right) \right).
\]

Then, the relative phase factor (RPF) between \( \rho(t_0) \) and \( \rho(t_1) \) along the curve \( C \) is the unitary matrix \( U \) which satisfies the equation

\[
W(t_1) = \hat{W}_1 U,
\]

where \( \hat{W}_1 \) satisfies \( \rho(t_1) = \pi(\hat{W}_1) \) and

\[
\hat{W}_1^* W(t_0) = W^*(t_0) \hat{W}_1.
\]

RPF is said to vanish when it is equal to the identity.

### 3.3. RPF for infinitesimal loop

The RPF for the infinitesimal loop

\[
(\theta^1, \ldots, \theta^i, \ldots, \theta^j + d\theta^j, \ldots, \theta^m) \leftrightarrow (\theta^1, \ldots, \theta^i + d\theta^i, \ldots, \theta^j + d\theta^j, \ldots, \theta^m)
\]

\[
\theta = (\theta^1, \ldots, \theta^i, \ldots, \theta^j, \ldots, \theta^m) \rightarrow (\theta^1, \ldots, \theta^i + d\theta^i, \ldots, \theta^j, \ldots, \theta^m)
\]

(16)
is calculated up to the second order of $d\theta$ by expanding the solution of the equation (15) to that order:

\begin{equation}
I + \frac{1}{2}W^{-1}F_{ij}W d\theta^i d\theta^j + o(d\theta)^2,
\end{equation}

\begin{equation}
F_{ij} = (\partial_i L^S_j - \partial_j L^S_i) - \frac{1}{2}[L^S_i, L^S_j].
\end{equation}

Note that $F_{ij}$ is a representation of the curvature form, and that RPF for any closed loop vanishes iff $F_{ij}$ is zero at any point in $\mathcal{M}$.

### 3.4. The SLD Cramér-Rao inequality

In parallel with the classical estimation theory, in the quantum estimation theory, we have the following SLD CR inequality, which is proved for the faithful state model by Helstrom [10, 11], for the pure state model by Fujiwara and Nagaoka [8], and for the general case by Fujiwara and Matsumoto [6]:

\begin{equation}
V_\theta[\hat{\theta}(\omega) | M] \geq (J^S(\theta))^{-1},
\end{equation}

i.e., $V_\theta[\hat{\theta}(\omega) | M] - (J^S(\theta))^{-1}$ is non-negative definite. Here $V_\theta[\hat{\theta}(\omega) | M]$ is a covariance matrix of an unbiased estimator $(\hat{\theta}, M, \Omega)$, and $J^S(\theta)$ is called SLD Fisher information matrix, and is defined by

\begin{equation}
J^S(\theta) \equiv \left[ \left\langle \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \theta^j} \right\rangle_{\rho(\theta)} \right] = \text{Re tr} \rho(\theta)L^S_i(\theta)L^S_j(\theta),
\end{equation}

which is nothing but the metric tensor of the inner product $\langle *, * \rangle$.

The inequality (18) is of special interest, because $J^{S-1}$, called SLD CR bound, is the one of the best bounds in the sense explained later.

To prove the inequality (18), we set some notations, and present some lemmas. For unbiased estimator $(\hat{\theta}, M, \Omega)$, we define the notation $M$ as follows,

\begin{equation}
M^i(\hat{\theta}, M, W) \equiv \int (\hat{\theta}^i(\omega) - \theta^i)M(d\omega)W.
\end{equation}

$Z[\hat{\theta}, M]$ is $m \times m$ matrix defined by

\begin{equation}
Z[\hat{\theta}, M] = \left[ \text{tr } M^i(\hat{\theta}, M, W) (M^j(\hat{\theta}, M, W))^* \right]
\end{equation}

**Lemma 4:** Following two inequalities are valid:

\begin{equation}
V[\hat{\theta} | M] \geq \text{Re } Z[\hat{\theta}, M].
\end{equation}
\[ V[\hat{\theta} | M] \geq Z[\hat{\theta}, M]. \]  

(22)

**Lemma 5:**

\[ \Re Z[\hat{\theta}, M] \geq J^{S-1} \]  

(23)

holds. The equality is valid iff

\[ M^j(\hat{\theta}, M, W) = \sum_k [J^{S-1}]^{j,k} L^j_k W, = \sum_k [J^{S-1}]^{j,k} M \left( \frac{h_W}{\partial \theta^k} \right), \]  

(24)

They are proved in almost the same manner as the strictly positive case (see p. 88 and p. 274 [12], respectively). Lemma 4 and 5 lead to the SLD CR inequality (18).

**Theorem 6:** SLD Fisher information gives a lower bound of covariance matrix of an unbiased measurement, i.e., (18) holds true.

The SLD CR bound (18) is the best bound in the following sense.

**Theorem 7:** Letting \( A \) be a real hermitian matrix which is larger than \( J^{S-1} \), that is, \( A > J^{S-1} \), there exists such an unbiased estimator \((\hat{\theta}, M, \Omega)\) that \( V[\hat{\theta} | M] \) is not smaller than \( A \).

**Proof:** Let \( v \in \mathbb{R}^m \) be the real vector such that

\[ \exists C \subset \mathbb{R} \quad \frac{\partial}{\partial \theta^i} \int_C \text{tr}(\rho(\theta)E_v dt) \neq 0, \quad (i = 1, \ldots, m), \]  

(25)

where \( E_v \) is a projection valued measure obtained by the spectral decomposition of \( \sum_{i,j} v_j (\theta^j I + [J^{S-1}]^{i,j} L^j_j) \), where \( I \) and \( v_i \) denotes identity and \( i \)th component of \( v \).

The condition (25) implies the existence of an estimator \( \hat{\theta}_v(\omega) \) which makes the triplet \((\hat{\theta}_v, E_v, \Omega)\) locally unbiased at \( \theta \). For that triplet \((\hat{\theta}_v, E_v, \Omega)\), we have

\[ v^T \left( V[\hat{\theta} | E_v] - J^{S-1} \right) v = 0. \]

If \( \varepsilon \) is small enough, for any real vector \( v \in \mathbb{R}^m \), \( v + \varepsilon v_0 \) satisfies the condition (25), or its equivalence,

\[ (v + \varepsilon v_0)^T \left( V[\hat{\theta}_{v+\varepsilon v_0} | E_{v+\varepsilon v_0}] - J^{S-1} \right) (v + \varepsilon v_0) = 0. \]

(26)

Let us assume that there exists a real matrix \( A \) which satisfies

\[ V[\hat{\theta} | M] \geq A > J^{S-1} \]
for any unbiased estimator. Then, by virtue of (26), we have for any real vector \( v_0 \in \mathbb{R}^m \) and enough small \( \varepsilon \),
\[
(v + \varepsilon v_0)^T (A - J^{S-1}) (v + \varepsilon v_0) = 0,
\]
whose second derivative with respect to \( \varepsilon \) yields
\[
v_0^T (A - J^{S-1}) v_0 = 0.
\]
Because \( v_0 \in \mathbb{R}^m \) is arbitrary, we have
\[
A - J^{S-1} = 0,
\]
which contradicts with the assumption \( A > J^{S-1} \). 

**Theorem 8:** If the model \( M \) has only one parameter, the equality in (18) is achievable.

**Proof:** Let \( E \) and \( \hat{\theta}(\ast) \) be a projection valued measurement and an estimator which satisfies,
\[
\left[ \int (\hat{\theta}(\omega) - \theta) M(d\omega) \right]^i = \sum_j [J^{S-1}]^{ji} L^S_j.
\]
Then, the triplet \( (\hat{\theta}, E, \Omega) \) is locally unbiased at \( \theta \) and attains SLD CR bound.

Theorem 8 implies the statistical significance of the natural metric \( \langle \ast, \ast \rangle \). A possible geometrical interpretation of (8): the closer two states \( \rho(t) \) and \( \rho(t + dt) \) are, the harder it is to distinguish \( \rho(t) \) from \( \rho(t + dt) \).

The SLD CR inequality (18) looks quite analogical to CR inequality in classical estimation theory. However, as will be found out later, the equality is not generally attainable.

### 3.5. The attainable Cramér-Rao type bound

In the previous section, SLD CR bound is proved to be the best bound. However, as will be turned out, this best bound is attainable only in the special cases, that is, the case when the model is locally quasi-classical. In general case, therefore, we must give up to find a tight lower bound of covariance matrix in the form of the matrix inequality. Instead, we determine the region \( V_\theta(M) \) of the map \( V[\ast] \) from unbiased estimators to \( m \times m \) real positive symmetric matrices (so far as no confusion is expected, we write \( \mathcal{V} \) for \( V_\theta(M) \)). Especially, the boundary \( bd\mathcal{V} \) is of interest, because \( \mathcal{V} \) is convex as in the following lemma.
**Lemma 9:** $\mathcal{V}$ is convex.

**Proof:** In this proof, we define the estimator to be the measurement which takes value in $\mathbb{R}^m$. Let $M$ and $M'$ be an unbiased estimator. Because

$$\lambda V[M] + (1 - \lambda)V[M'] = V[\lambda M + (1 - \lambda)M']$$

holds true and $\lambda M + (1 - \lambda)M'$ is an unbiased estimator, we have the lemma. 

**Lemma 10:** If a matrix $V$ is a member of $\mathcal{V}$, $V + V_0$ is also a member of $\mathcal{V}$ for any arbitrary nonnegative real symmetric matrix $V_0$.

**Proof:** Let $(\hat{\theta}, M, \Omega)$ be a locally unbiased estimator whose covariance matrix is $V$, and define $\varepsilon(\omega)$ by

$$\varepsilon(\omega) \equiv V_0^{1/2}(V[\hat{\theta} | M])^{-1/2} (\hat{\theta}(\omega) - \theta).$$

Then, $\hat{\theta}^*(\omega) = \hat{\theta}(\omega) + \varepsilon(\omega)$ is also a locally unbiased estimator, and its covariance matrix is equal to $V + V_0$. 

To obtain $bd\mathcal{V}$, the following procedure is used in this thesis. Define an inner product of two real symmetric matrices $A$ and $B$ by $\text{Tr} AB$. Then, the set $\{V | \text{Tr} VG = \text{const.} \}$ is a hyperplain perpendicular to the vector $G$. Because of lemmas 9-10, $bd\mathcal{V}$ is the collection of all the matrices $V \in \mathcal{V}$ which achieve the minimum of $\text{Tr} GV$ for a certain symmetric real nonnegative definite matrix $G$.

However, when the model has too many parameters, the dimension of the space $\text{Sym}(m)$ of real symmetric matrices is so large that $\mathcal{V}_\theta(M)$ is extremely hard to determine. In such cases, we calculate

$$\text{CR}(G, \theta, M) \equiv \min \{ \text{Tr} GV | V \in \mathcal{V}_\theta(M) \}$$

for an arbitrary nonnegative symmetric real matrix $G$, and call it the attainable CR (Cramér-Rao) type bound. The matrix $G$ is called weight matrix. Often, we drop $G$, $\theta$ and/or $M$ when no confusion is expected. If $\text{CR}(G, \theta, M)$ is smaller than $\text{CR}(G, \theta', M')$ for any weight matrix $G$, the $\mathcal{V}_\theta(M)$ of is located in the ‘lower part’ of $\text{Sym}(m)$ compared with that of $\mathcal{V}_{\theta'}(M')$.

To make the estimational meaning of (27) clear, let us consider a diagonal weight matrix $G = \text{diag}(g_1, g_2, ..., g_m)$. Letting $v_{ii}$ be the $(i,i)$-th component of $V[M]$,

$$\text{Tr} GV[M] = \sum_i g_i v_{ii},$$
is the weighed sum of the variances of the estimations of the parameters $\theta^i$ $(i = 1, ..., m)$. If the accuracy of estimation of, for example, the parameter $\theta^1$ is required more than other parameters, then $g_1$ is set larger than any other $g_i$, and the estimator which minimize $\sum_i g_i v_{ii}$ is to be used.

3.6. The locally quasi-classical model and the quasi-classical model

In this section, the condition for SLD CR bound to be attainable is reviewed briefly, in the case of the faithful model, any member of which is faithful, i.e., a reversible operator. We denote the space of the faithful states by $\mathcal{P}_+(\mathcal{H})$, or simply by $\mathcal{P}_+$. For mathematical simplicity, the dimension $d$ of the Hilbert space $\mathcal{H}$ is assumed to be finite. The author believes the essence of the discussion will not be damaged by this restriction.

As for the equality in (18) in the faithful model, we have the following theorem, which is proved by Nagaoka [18].

**Theorem 11:** The equality in (18) is attainable at $\theta$ iff $[L^S_i(\theta), L^S_j(\theta)] = 0$ for any $i, j$. Letting $|\omega\rangle$ be a simultaneous eigenvector of the matrices $\{L^S_j(\theta)|j = 1, ..., m\}$ and $\lambda_i(\omega)$ be the eigenvalue of $L^S_i(\theta)$ corresponding to $|\omega\rangle$, the equality is attained by the estimator $(\hat{\theta}(\theta), M(\theta), \Omega)$ such that

$$\Omega = \{\omega|\omega = 1, ..., n\},$$

$$M(\theta)(\omega) = |\omega\rangle\langle\omega|,$$

$$\hat{\theta}_i(\theta)(\omega) = \theta^i + \sum_{k=1}^{d} [(J^S)^{-1}]^{jk} \lambda_k(\omega).$$

(28)

The model $\mathcal{M}$ is said to be locally quasi-classical at $\theta$ iff $L^S_i(\theta)$ and $L^S_j(\theta)$ commute for any $i, j$, because in this case (18) gives the attainable lower bound of the covariance matrix of the unbiased estimator, as its classical counterpart does. However, it should be noted that even if the model is locally quasi-classical at any $\theta \in \Theta$, theorem 11 do not tell us the optimal experiment scheme, because the measurement $M(\theta)$ in (28) generally depends on the unknown parameter $\theta$ (so does the best experiment scheme).

Therefore, let us move to easier case. Suppose that $L^S_i(\theta)$ and $L^S_j(\theta')$ commute even when $\theta \neq \theta'$, in addition to being locally quasi-classical at any $\theta \in \Theta$. Then, the measurement $M(\theta)$ in (28), denoted by $M_{\text{best}}$ in the remainder of the section, is uniformly optimal for all $\theta$ (so is the corresponding scheme). We say such a model is quasi-classical [28].
After the best experiment is done, the rest of our task is to estimate the value of the parameter $\theta$, where $M_{\text{best}}$ is substituted into $M$. Hence, in this case, the quantum estimation reduces to the classical estimation.

4. Uhlmann Connection and the Estimation Theory

4.1. Some new facts about RPF

In this subsection, we derive conditions for RPF to vanish, which is used to characterize the classes of manifold defined in the previous subsection. For notational simplicity, the argument $\theta$ is omitted, as long as the omission is not misleading.

Theorem 12: RPF for any closed loop vanishes iff $\left[ L^S_i(\theta), L^S_j(\theta) \right] = 0$ for any $\theta \in \Theta$. In other words,

$$F_{ij}(\theta) = 0 \iff \left[ L^S_i(\theta), L^S_j(\theta) \right] = 0.$$  \hfill (29)

Proof: If $F_{ij}$ equals zero, then both of the two terms in the left-hand side of (17) must vanish, because the first term is Hermitian and the second term is skew Hermitian. Hence, if $F_{ij} = 0$, $[L^S_i, L^S_j]$ vanishes.

On the other hand, the identity $\partial_i \partial_j \rho - \partial_j \partial_i \rho = 0$, or its equivalence

$$\left( \partial_i L^S_j - \partial_j L^S_i - \frac{1}{2} [L^S_i, L^S_j] \right) \rho + \rho \left( \partial_i L^S_j - \partial_j L^S_i + \frac{1}{2} [L^S_i, L^S_j] \right) = 0,$$

implies that $\partial_i L^S_j - \partial_j L^S_i$ vanishes if $[L^S_i, L^S_j] = 0$, because $\partial_i L^S_j - \partial_j L^S_i$ is Hermitian and $\rho$ is positive definite. Thus we see $F_{ij} = 0$ if $L^S_i$ and $L^S_j$ commute.

A manifold $M$ is said to be parallel when the RPF between any two points along any curve vanishes. From the definition, if $M$ is parallel, RPF along any closed loop vanishes, but the reverse is not necessarily true. The following theorem is a generalization of Uhlmann’s theory of $\Omega$-horizontal real plane [26].

Theorem 13: The following three conditions are equivalent.

(i) $M$ is parallel.

(ii) Any element $\rho(\theta)$ of $M$ writes $M(\theta) \rho_0 M(\theta)$, where $M(\theta)$ is Hermitian and $M(\theta_0) M(\theta_1) = M(\theta_1) M(\theta_0)$ for any $\theta_0, \theta_1 \in \Theta$.

(iii) $\forall i, j, \forall \theta_0, \theta_1 \in \Theta, \left[ L^S_i(\theta_0), L^S_j(\theta_1) \right] = 0$.

Proof: Let $W(\theta_t) = M(\theta_t)W_0$ be a horizontal lift of $\{\rho(\theta_t) | t \in \mathbb{R}\} \subset M$. Then, $W^*_0 W(\theta_t) = W^*(\theta_t)W_0$ implies $M(\theta_t) = M^*(\theta_t)$,
and \( W^*(\theta_{t_0}) W(\theta_{t_1}) = W^*(\theta_{t_1}) W(\theta_{t_0}) \) implies \( M(\theta_{t_0}) M(\theta_{t_1}) = M(\theta_{t_1}) M(\theta_{t_0}) \). Thus we get (1) \( \Rightarrow \) (2). Obviously, the reverse also holds true. For the proof of (2) \( \Leftrightarrow \) (3), see pp. 31–33 [28]. □

4.2. **Uhlmann’s parallelism in the quantum estimation theory**

To conclude the section, we present the theorems which geometrically characterize the locally quasi-classical model and quasi-classical model, by the vanishing conditions of RPF, implying the close tie between Uhlmann parallel transport and the quantum estimation theory. They are straightforward consequences of the definitions of the terminologies and theorems 11-13.

**Theorem 14:** \( \mathcal{M} \) is locally quasi-classical at \( \theta \) iff \( F_{ij}(\theta) = 0 \) for any \( i, j \).

\( \mathcal{M} \) is locally quasi-classical at any \( \theta \in \Theta \) iff the RPF for any loop vanishes.

**Theorem 15:** \( \mathcal{M} \) is quasi-classical iff \( \mathcal{M} \) is parallel.

5. **The Pure State Estimation Theory**

5.1. **Historical review of the theory and the purpose of the section**

First, we review the history of quantum estimation theory to clarify the purpose of the section.

In parallel with the classical estimation theory, in 1967, Helstrom show that in the faithful state model, the covariance matrix is larger than or equal to the inverse of SLD Fisher information matrix, and that in the one parameter faithful model, the bound is attainable [10, 11].

On the other hand, in the multi-parameter model, it is proved that there is no matrix which makes attainable lower bound of covariance matrix, because of non-commutative nature of quantum theory. Therefore, we deal with the attainable CR type bound defined in the subsection 3.5. Lower bounds of \( \text{Tr} G V_{\theta}[M] \) is, attainable or not, called *Cramér-Rao (CR) type bound*.  

Five years after Helstrom’s work, Yuen and Lax found out the exact form of the attainable CR type bound of the Gaussian state model, which is a faithful 2-parameter model obtained by superposition of coherent states by Gaussian kernel [27]. Their work is remarkable not only because it was first calculation of the attainable CR type bound of a multi-parameter model, but also because they established a kind of methodology, which we
call *indirect approach* hereafter, to calculate the attainable CR type bound. First an auxiliary bound which is not generally attainable is found out and then it is proved to be attained in the specific cases. In their work, they used so-called RLD bound, which was used and generalized by several authors.

Holevo completed their work by solving analytically subtle problems and generalizing RLD bound [12]. In 1998, Nagaoka calculated the attainable CR type bound of the faithful 2-parameter spin-1/2 model using Nagaoka bound, which is also another auxiliary bound [15].

The models had been assumed to be faithful till Fujiwara and Nagaoka formulated the problem in the pure state model, and calculated the CR type bound for the 1-parameter model by use of generalized SLD bound and that of the 2-parameter coherent model by use of generalized RLD bound in 1995 [8, 9].

The approach in this section, called *direct approach* in contrast with *indirect approach*, is essentially different from the approaches of other authors. We reduce the given minimization problem to the problem which is easy enough to be solved directly by elementary calculus. The methodology is successfully applied to the general 2-parameter pure state model, coherent model with arbitrary number of parameters, and the minimization of $\text{Tr} J^S(\theta)V_\theta[M]$ for arbitrary pure state model. These are relatively general category in comparison with the cases treated by other authors. In the 2-parameter pure state model, the existence of the order parameter $\beta$ which is a good index of noncommutative nature between the parameters.

As a by-product, we have remarkable corollary, which asserts that even for non-commutative cases, a simple measurement attains the lower bound.

### 5.2. Notations

In this subsection, we consider data space $\Omega$ to be $\mathbb{R}^m$, and each $\omega \in \Omega$ to be the estimate $\hat{\theta}$.

The pure state model $\mathcal{M}$ is assumed to be induced by $\mathcal{M} = \pi(\mathcal{N})$ from the manifold

$$\mathcal{N} = \{ |\phi(\theta)\rangle \ | \phi(\theta) \in \tilde{\mathcal{H}} \},$$

where $\tilde{\mathcal{H}}$ is the set of the members of $\mathcal{H}$ with unit length,

$$\tilde{\mathcal{H}} = \{ |\phi\rangle \ | \phi \in \mathcal{H}, \langle\phi|\phi\rangle = 1 \}.$$

If this assumption is made, the horizontal lift of the tangent vector is taken for granted. Uniqueness is proved mostly in the same way as the proof of theorem 3.
We denote by $|l_X\rangle$ the matrix representation of the horizontal lift of $X \in T_{\rho(\theta)}(M)$, and $|l_i\rangle$ is short for $|l_{\partial/\partial \theta^i}\rangle$. $|l_X\rangle$ satisfies
\[X\rho(\theta) = \frac{1}{2}|l_X\rangle\langle \phi(\theta)| + |\phi(\theta)\rangle\langle l_X|, \tag{30}\]
and
\[\langle l_X|\phi(\theta)\rangle = 0. \tag{31}\]

Notice that $\text{span}_R\{ |l_i\rangle \mid i = 1, \ldots, m \}$ is a representation of $T_{\rho(\theta)}(M)$ because of unique existence of $|l_X\rangle$. Therefore, we often also call the matrix representation $|l_X\rangle$ horizontal lift.

We call $M^i(M, |\phi(\theta)\rangle)$ estimation vector of the parameter $\theta^i$ by a measurement $M$ at $|\phi(\theta)\rangle$. An estimation vector $M^i(M, |\phi(\theta)\rangle)$ is said to be locally unbiased iff $M$ is locally unbiased. The local unbiasedness conditions for estimating vectors writes
\[\langle \phi(\theta)|M^i(M, |\phi(\theta)\rangle) = 0, \]
\[\text{Re}\langle l_j(\theta)|M^i(M, |\phi(\theta)\rangle) = \delta^i_j (i, j = 1, \ldots, m). \tag{32}\]

Often, we omit the argument $\theta$ of $|l_j(\theta)\rangle, |\phi(\theta)\rangle, \rho(\theta)$, and $J^S(\theta)$ and denote them simply by $|l_j\rangle, |\phi\rangle, \rho, J^S$.

We denote the ordered pair of vectors $\left[ |l_1\rangle, |l_2\rangle, \ldots, |l_m\rangle \right]$ by $L$. In this notation, the SLD Fisher information matrix $J^S$ writes
\[J^S = \text{Re}\ast L \equiv \text{Re} \left[ \langle l_i|l_j\rangle \right],\]
and the imaginary part of $L^\ast L$ is denoted by $\tilde{J}$. Generally, for the ordered pairs $X = [ |x^1\rangle, |x^2\rangle, \ldots, |x^m\rangle ]$, $Y = [ |y^1\rangle, |y^2\rangle, \ldots, |y^m\rangle ]$, of vectors, we define
\[X^\ast Y = \left[ \langle x^i|y^j\rangle \right]\]
for notational simplicity.

Then, letting $X$ be
\[\left[ M^1(M, |\phi\rangle), M^2(M, |\phi\rangle), \ldots, M^m(M, |\phi\rangle) \right],\]
the unbiasedness conditions (32) writes
\[\text{Re} X^\ast L = I_m, \quad \text{Re} \{ (M^i(M, |\phi\rangle))^\ast |l^i\rangle \} = I_m, \tag{33}\]
where $I_m$ is the $m \times m$ unit matrix, and the matrix $Z_\theta[M]$ defined in the equation (20) writes

$$ Z_\theta[M] = X^*X $$

5.3. The commuting theorem and the locally quasi-classical model

In this subsection, the necessary and sufficient condition for SLD CR bound to be attainable is studied. Fujiwara proved the following theorem [7].

**Theorem 16:** (Fujiwara [7]) SLD CR bound is attainable iff SLD’s $\{L_i^S|i = 1,\ldots,m\}$ can be chosen so that

$$ [L_i^S, L_j^S] = 0, \ (i, j = 1,\ldots,m). $$

We prove another necessary and sufficient condition which is much easier to check for the concrete examples, by use of the following commuting theorem, which plays key role in our direct approach to pure state estimation.

**Theorem 17:** If there exists an unbiased measurement $M$ such that

$$ V[M] = \text{Re}X^*X, \quad (34) $$

where the ordered pair $X$ is

$$ X = [M^1(M, |\phi\rangle), M^2(M, |\phi\rangle), \ldots, M^m(M, |\phi\rangle)], \quad (35) $$

then,

$$ \text{Im}X^*X = 0 \quad (36) $$

holds true. Conversely, if (36) holds true for some ordered pair $X$ of vectors, then there exists a simple, or projection valued, unbiased estimator $E$ which satisfies (34), (35), and

$$ E(\{\hat{\theta}_\kappa\})^2 = E(\{\hat{\theta}_0\}) = E_0, \ E\left(\mathbb{R}^m / \bigcup_{\kappa=0}^{m} \{\hat{\theta}_\kappa\}\right) = 0, \quad (37) $$

for some $\{\hat{\theta}_\kappa|\hat{\theta}_\kappa \in \mathbb{R}^m, \kappa = 0,\ldots,m+1\}$, where $E_0$ is a projection onto orthogonal complement subspace of $\text{span}_C\{X\}$.

**Proof:** If (34) holds, inequality (22) in lemma 4 leads to

$$ \text{Re}X^*X \geq X^*X, \quad \text{or} \quad 0 \geq \text{Im}X^*X, $$
which implies \( \text{Im}X^*X = 0 \).

Conversely, let us assume that (36) holds true. Applying Schmidt’s orthogonalization to the system \( \{|\phi\rangle, X\} \) of vectors, we obtain the orthonormal system \( \{|b^i\rangle \mid i = 1, \ldots, m + 1\} \) by which the system \( X \) of vectors writes of vectors such that,

\[
X = \left[ \sum_{j=1}^{m+1} \lambda_j^1 |b^j\rangle, \sum_{j=1}^{m+1} \lambda_j^2 |b^j\rangle, \ldots, \sum_{j=1}^{m+1} \lambda_j^m |b^j\rangle \right],
\]

where \( \lambda_j^i (i = 1, \ldots, m, j = 1, \ldots, m + 1) \) are real numbers. Letting \( O = [o_{ij}] \) be a \((m + 1) \times (m + 1)\) real orthogonal matrix such that

\[
\langle \phi | \sum_{j=1}^{m+1} o_{ij}^i |b^j\rangle \neq 0,
\]

and denoting \( \sum_{j=1}^{m+1} o_{ij}^i |b^j\rangle \) by \( |b^i\rangle \), \( i \)th member of the ordered pair \( X \) writes

\[
\sum_{j=1}^{m+1} \lambda_j^i \sum_{k=1}^{m+1} o_{j}^k |b^k\rangle = \sum_{k=1}^{m+1} \left( \sum_{j=1}^{m+1} \lambda_j^i o_{j}^k \right) |b^k\rangle = \sum_{k=1}^{m+1} \sum_{j=1}^{m+1} \lambda_j^i o_{j}^k |b^k\rangle \langle b^k | \phi \rangle.
\]

Therefore, since the system \( \{|b^i\rangle \mid i = 1, \ldots, m + 1\} \) of vectors is orthonormal, we obtain an unbiased measurement which satisfies (37) as follows:

\[
\hat{\theta}_\kappa = \frac{\sum_{j=1}^{m+1} \lambda_j^i o_{j}^\kappa}{\langle b^\kappa | \phi \rangle}, \quad \kappa = 1, \ldots, m + 1, \quad \hat{\theta}_0 = 0,
\]

\[
E(\{\hat{\theta}_\kappa\}) = |b^\kappa\rangle \langle b^\kappa |, \quad \kappa = 1, \ldots, m + 1 \quad E(\hat{\theta}_0) = I - \sum_{\kappa=1}^{m+1} |b^\kappa\rangle \langle b^\kappa |.
\]

Here, \( I \) is the identity in \( \mathcal{H} \).

**Theorem 18:** SLD CR bound is attainable iff

\[
\tilde{J} = \text{Im}L^*L = 0
\]

\( \langle l_i | l_j \rangle \) is real for any \( i, j \). Conversely, if (38) holds true, SLD CR bound is achieved by a simple measurement, i.e., a projection valued measurement.

**Proof:** If SLD CR bound is attainable, by virtue of lemma 4-5, we have (36) and (24), which lead directly to (38).

Conversely, if \( \text{Im}\langle l_j | l_k \rangle = 0 \) for any \( j, k \), by virtue of commuting theorem, there exists such a simple measurement \( E \) that

\[
\sum_k [J^{S-1}]^{i,k} |l_k\rangle = M(E, |\phi\rangle).
\]
Elementary calculations show that the covariance matrix of this measurement equals $J S^{-1}$.

Our theorem is equivalent to Fujiwara’s one, because by virtue of commuting theorem, $\langle l_j | l_i \rangle$ is real iff there exist such SLD’s that $L_i^S$ and $L_j^S$ commute for any $i, j$. However, our condition is much easier to be checked, because SLD’s are not unique in the pure state model.

**Example 19:** Often, a model is defined by an initial state and generators,

$$\rho(\theta_0) = \rho, \quad \partial_\theta \rho(\theta) = i[H_i(\theta), \rho(\theta)], \quad \rho(\theta) = \pi(|\phi(\theta)\rangle).$$

Then, $\langle l_j | l_i \rangle$ is real iff $\langle \phi(\theta) | [H_i(\theta), H_j(\theta)] |\phi(\theta)\rangle$ is 0. Because of theorem 17, $\langle \phi(\theta) | [H_i(\theta), H_j(\theta)] |\phi(\theta)\rangle = 0$ is equivalent to the existence of generators which commute with each other, $[H_i(\theta), H_j(\theta)] = 0$.

Because of this example and the theorem by Fujiwara, we may metaphorically say that SLD CR bound is attainable iff any two parameter has ‘classical nature’ at $\theta$, because often classical limit of a quantum system is obtained by taking such a limit that commutation relations of observables tend to 0. Throughout the paper, we say that a manifold $\mathcal{M}$ is locally quasi-classical at $\theta$ iff $\langle l_j | l_i \rangle$ is real at $\theta$. The following remark describes another ‘classical’ aspect of the condition $\text{Im} \langle l_j | l_i \rangle = 0$.

**Example 20:** The model $\mathcal{M} = \pi(\mathcal{N})$, where $\mathcal{N}$ is a real span of some orthonormal basis of $\mathcal{H}$, is locally quasi-classical at any point.

As is illustrated in this example, when the model $\mathcal{M} = \pi(\mathcal{N})$ is locally quasi-classical at $\theta_0$, $\mathcal{N}$ behaves like an element of real Hilbert space around $\theta_0$. Metaphorically speaking, $|\phi(\theta)\rangle$’s phase parts don’t change around $\theta_0$ at all, and $\mathcal{N}$ looks like the family of the square root of the probability distributions.

### 5.4. The reduction theorem and the direct approach

**Theorem 21:** (Naimark’s theorem, see pp.64–68 [12].) Any generalized measurement $M$ in $\mathcal{H}$ can be dilated to a simple measurement $E$ in a larger Hilbert space $\mathcal{K} \subset \mathcal{H}$, so that

$$M(B) = P E(B) P$$

will hold, where $P$ is the projection from $\mathcal{K}$ onto $\mathcal{H}$. 

Naimark's theorem, mixed with commuting theorem, leads to the following reduction theorem, which sits at the heart of our direct approach.

**Theorem 22:** Let \( M \) be a \( m \)-dimensional manifold in \( \mathcal{P}_1 \), and \( B_\theta \) be a system \( \{ |\phi^i\rangle, |l_i^j\rangle \mid i = 1, \ldots, m \} \) of vectors in \( 2m + 1 \)-dimensional Hilbert space \( \mathcal{K}_\theta \) such that

\[
\langle \phi^i | l_j^i \rangle = \langle \phi^i | l_j^i \rangle = 0, \quad \langle \phi^i | \phi^i \rangle = \langle \phi | \phi \rangle = 1
\]

for any \( i, j \). Then, for any locally unbiased estimator \( M \) at \( \theta \) in \( \mathcal{H} \), there is a simple 'locally unbiased' measurement \( E \) in \( \mathcal{K}_\theta \),

\[
|x^i\rangle = \mathbf{M}^i(E, |\phi^i\rangle) \in \mathcal{K}_\theta
\]

\[
\langle x^i | \phi^j \rangle = 0,
\]

\[
\operatorname{Re} \langle x^i | l^j_i \rangle = \delta^i_j (i, j = 1, \ldots, m),
\]

whose 'covariance matrix' \( V[E] \) equals \( V[M] \),

\[
V[M] = V[E] \equiv \left[ \int (\hat{\theta}^i - \theta^i)(\hat{\theta}^j - \theta^j)\langle \phi^i | E(d\hat{\theta})|\phi^j \rangle \right],
\]

**Proof:** For any locally unbiased measurement \( M \), there exists a Hilbert space \( \mathcal{H}_M \) and a simple measurement \( E_M \) in \( \mathcal{H}_M \) which satisfies (39) by virtue of Naimark’s theorem. Note that \( E_M \) is also locally unbiased. Mapping \( \text{span}_\mathbb{C} \{ |\phi\rangle, |l_i\rangle, \mathbf{M}^i(M, |\phi\rangle) \mid i = 1, \ldots, m \} \) isometrically onto \( \mathcal{K}_\theta \) so that \( \{ |\phi\rangle, |l_i\rangle \mid i = 1, \ldots, m \} \) are mapped to \( \{ |\phi^i\rangle, |l^i_j\rangle \mid i = 1, \ldots, m \} \), we denote the images of \( \{ \mathbf{M}^i(M, |\phi\rangle) \mid i = 1, \ldots, m \} \) by \( \{ |x^i\rangle \mid i = 1, \ldots, m \} \).

Then, by virtue of the commuting theorem, we can construct a simple measurement \( E \) in \( \mathcal{K}_\theta \) satisfying the equations (40) - (42).

In our direct approach, the reduction theorem reduces the determination of \( \mathcal{V} \) to the determination of the set of matrices

\[
V = \operatorname{Re}X^*X
\]

where a system \( X = [|x^1\rangle, |x^2\rangle, \ldots, |x^m\rangle] \) of elements of \( \mathcal{K}_\theta \ominus \{ |\phi^i\rangle \} \) which satisfies (36) and (41). In the same way, we minimize

\[
\operatorname{Tr} G \operatorname{Re}X^*X,
\]

where \( \{ |x^i\rangle \mid i = 1, \ldots, m \} \in \mathcal{K}_\theta \ominus \{ |\phi^i\rangle \} \), under the restriction such that the equations (36) and (41) are satisfied, instead of minimization of \( \operatorname{tr} GV \) where \( V \) runs through \( \mathcal{V} \).

Now, the problem is simplified to the large extent, because we only need to treat with vectors \( \{ |x^i\rangle \mid i = 1, \ldots, m \} \) in finite dimensional Hilbert space \( \mathcal{K}_\theta \) instead of measurements, or operator valued measures.
We conclude this subsection with a corollary of reduction theorem, which is seemingly paradoxical, since historically, non-projection-valued measurement is introduced to describe measurements of non-commuting observables.

**Corollary 23:** When the dimension of $\mathcal{H}$ is larger than or equal to $2m+1$, for any unbiased measurement $M$ in $\mathcal{H}$, there is a simple measurement $E$ in $\mathcal{H}$ which has the same covariance matrix as that of $M$.

**Proof:** Choose $\{|l'_i\rangle | i = 1, ..., m\}$ to be $\{|l_i\rangle | i = 1, ..., m\}$. Especially, if $\mathcal{H}$ is infinite dimensional, as is the space of wave functions, the assumption of the corollary is always satisfied.

5.5. **Lagrange’s method of indeterminate coefficients in the pure state estimation theory**

Now, we apply our direct approach to the problems presented in the end of subsection 3.5. To minimize the functional $\text{Tr} G \text{Re} X^* X$ of vectors in $\mathcal{K}_\theta$, Lagrange’s indeterminate coefficients method is employed. First, denoting an ordered pair $\{|l'_i\rangle | i = 1, ..., \}$ of vectors in $\mathcal{K}_\theta$ also by $L$, the symbol which is used also for an ordered pair $\{|l_i\rangle | i = 1, ..., \}$ of vectors in $\mathcal{H}$, we define a function $\text{Lag}(X)$ by

$$\text{Lag}(X) \equiv \text{Re} \text{Tr} X^* X G - 2 \text{Tr} ((\text{Re} X^* L - I_m) \Xi) - \text{Tr} \text{Im} X^* X \Lambda,$$

(43)

where $\Xi, \Lambda$ are matrices whose components are Lagrange’s indeterminate coefficients. Here, $\Lambda$ can be chosen to be antisymmetric, for

$$\text{Tr} \text{Im} X^* X \Lambda = \text{Tr} \text{Im} X^* X (\Lambda - \Lambda^T)/2$$

holds true and only skew symmetric part of $\Lambda$ appears in (43).

From here, we follow the routine of Lagrange’s method of indeterminate coefficients. Differentiating $L(X + \varepsilon \delta X)$ with respect to $\varepsilon$ and substituting 0 into $\varepsilon$ in the derivative, we get

$$\text{Re} \text{Tr} (\delta X^* (2XG - 2L\Xi - 2iX\Lambda)) = 0.$$ 

Because $\delta X$ is arbitrary,

$$X(G - i\Lambda) = L\Xi$$

(44)

is induced.

Multiplying $X^*$ to both sides of (44), the real part of the outcoming equation, together with (33), yields

$$\Xi = \text{Re} X^* XG = VG.$$  

(45)
Substituting (45) into (44), we obtain
\[ X(G - i\Lambda) = LVG. \] (46)

If (33), (36), (46) and \( V = \text{Re}X^*X \) are solved for \( V \), \( X \) and real skew symmetric matrix \( \Lambda \), our problems will be perfectly solved. However, so far, solutions only for special cases are known.

The rest of this subsection is devoted to the proof of the theorem which claim a little stronger assertion than the corollary 23.

Though the real linear space \( \text{span}_R\{L\} \) is always \( m \)-dimensional for the parameters not to be redundant, the dimension of the complex linear space \( \text{span}_C\{L\} \), or the rank of \( L \), is not necessarily equal to \( m \). If \( \text{rank}_C L = \text{rank} G = m \) is assumed, since the rank of the matrix \( V \) is \( m \) as is proved soon, the rank of the left hand side of (46) is \( m \), so is the rank of the right hand side, implying that \( G - i\Lambda \) is invertible. The rank of the matrix \( V \) is \( m \) because (33) implies that the dimension of \( \text{span}_R\{X\} \) is \( m \).

Since \( X \) is given by \( LVG(G - i\Lambda)^{-1} \), we can conclude that \( \text{span}_C\{X\} \) should be a subspace of \( \text{span}_C\{L\} \). Therefore, by the same argument as in the proof of the corollary 23, we obtain the following theorem.

**Theorem 24:** Suppose that the dimension \( \mathcal{H} \) is larger than or equal to \( m + 1 \), and that the dimension of the complex linear space \( \text{span}_C\{L\} \) is \( m \). Then, for any strictly positive weight matrix \( G \), the attainable CR type bound \( \text{CR}(G) \) is attained by a simple measurement.

### 5.6. The model with two parameters

In this subsection, we determine \( V \) for the arbitrary 2-parameter pure state model.

The equation (46), mixed with (36), leads to
\[ (G - i\Lambda)V(G - i\Lambda) = GVL^*LVG. \] (47)

whose real part and imaginary part are
\[ GVG - AV\Lambda = GVJ^S VG, \] (48)

and
\[ GV\Lambda + AVG = -GV\tilde{J}VG, \] (49)

where \( \tilde{J} \) denotes \( \text{Im}L^*L \), respectively.

We assert that when the matrix \( G \) is strictly positive, (47) is equivalent to the existence of \( X \) which satisfies (33), (36), (47), and \( V = \text{Re}X^*X \). If real
positive symmetric matrix $V$ and real antisymmetric matrix $\Lambda$ satisfying (47) exist, $X$ which satisfies (46) and (36) is given by $X = UV^{1/2}$, where $U$ is a $2m + 1$ by $m$ complex matrix such that $U^*U = I_m$. If $G$ is strictly positive, $X = UV^{1/2}$ also satisfies (33), because

$$VG = \text{Re}X^*LVG$$

is obtained by multiplication of $X^*$ to and taking real part of the both sides of (46), and our assertion is proved.

Hence, our task is to solve (48) and (49) for real positive symmetric matrix $V$ and real antisymmetric matrix $\Lambda$, if $G$ is strictly positive. When $G$ is not strictly positive, after solving (48) and (49), we must check whether there exists an ordered pair $X$ of vectors which satisfies (33), (36) and $V = \text{Re}X^*X$.

In the remainder of this subsection, we use the coordinate system where $J^S$ is equal to the identity $I_m$. Given an arbitrary coordinate system $\{\theta^i| i = 1, ..., m\}$, such a coordinate system $\{\theta'^i| i = 1, ..., m\}$ is obtained by the following coordinate transform:

$$\theta'^i = \sum_{j=1}^{m} [(J^S)^{1/2}]_{ij} \theta^j \ (i = 1, ..., m) \ (50)$$

By this coordinate transform, $V$ is transformed as:

$$V' = (J^S)^{1/2}V(J^S)^{1/2}. \ (51)$$

Therefore, the result in the originally given coordinate is obtained as a transformation of the result in the coordinate system $\{\theta'^i| i = 1, ..., m\}$, by using (51) in the converse way.

So far, we have not assumed dim $\mathcal{M} = 2$. When dim $\mathcal{M} = 2$, covariance matrices are members of the space $Sym(2)$ of $2 \times 2$ symmetric matrices which is parameterized by the real variables $x, y$, and $z$, where

$$Sym(2) = \left\{ V \left| V = \begin{bmatrix} z+x & y \\ y & z-x \end{bmatrix} \right. \right\}.$$

Before tackling the equations (48) and (49), three useful facts about this parameterization are noted. First, letting $A$ is a symmetric real matrix which is represented by $(A_x, A_y, A_z)$ in the $(x, y, z)$-space, the set $\mathcal{P}\mathcal{M}(A)$ of all matrices larger than $A$ is

$$\mathcal{P}\mathcal{M}(A) = \{(x, y, z) \mid (z - A_z)^2 - (x - A_x)^2 - (y - A_y)^2 \geq 0, (z - A_z) \geq 0\},$$

that is, interior of a upside-down corn with its vertex at $A = (A_x, A_y, A_z)$. Hence, $V$ is a subset of $\mathcal{P}\mathcal{M}(I_m)$, or inside of an upside-down corn with its
vertex at (0, 0, 1) because of SLD CR bound. When the model $\mathcal{M}$ is locally quasi-classical at $\theta$, $\mathcal{V}$ coincides with $\mathcal{P}\mathcal{M}(I_m)$.

Second, an action of rotation matrix $R_\theta$ to $V$ such that $R_\theta V R_\theta^T$, where

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix},$$

corresponds to the rotation in the $(x, y, z)$-space around $z$-axis by the angle $2\theta$.

Third, we have the following lemma.

**Lemma 25:** $\mathcal{V}$ is rotationally symmetric around $z$-axis, if the coordinate in $\mathcal{M}$ is chosen so that $J^S$ writes the unit matrix $I_m$.

**Proof:** $\mathcal{V}$ is the set of every matrix which writes $\text{Re}X^*X$ using a $2m + 1$ by $m$ complex matrix $X$ satisfying (33) and (36). Therefore, the rotational symmetry of $\mathcal{V}$ around $z$-axis is equivalent to the existence of a $2m + 1 \times m$ complex matrix $Y$ satisfying (33), (36) and

$$\forall \theta \ R_\theta X^* X R_\theta^T = Y^* Y, \quad (52)$$

for any $2m + 1 \times m$ complex matrix $X$ which satisfies (33) and (36). Because of $L^* L = I_m + i\tilde{J}$, elementary calculation shows that

$$L^* L = \text{Re} R_\theta L^* L R_\theta^T,$$

or, that for some unitary matrix $U$ in $K_\theta$,

$$L^* U = R_\theta L^*,$$

which leads, together with (33), to

$$\text{Re} L^* U X = R_\theta. \quad (53)$$

Therefore,

$$Y = U X R_\theta^T$$

satisfies (52), and we have the lemma.

Because of lemma 25, $\mathcal{V}$ is determined if the boundary of the intersection $\tilde{\mathcal{V}}$ of $\mathcal{V}$ and the $zx$-plane is calculated. Note that the ‘inner product’ $\text{Tr} G V$ of $G$ and $V$ does not take its minimum at $V \in \tilde{\mathcal{V}}$ unless $G$ is in the $zx$-plane. Therefore, to obtain $\text{bd} \tilde{\mathcal{V}}$, only diagonal weight matrix is needed to be considered.
Let us begin with the case of a positive definite weight matrix. In this case, we only need to deal with (48) and (49). Let
\[
\tilde{J} = \begin{bmatrix} 0 & -\beta \\ \beta & 0 \end{bmatrix}, \quad \text{and} \quad \Lambda = \begin{bmatrix} 0 & -\lambda \\ \lambda & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & g \end{bmatrix}, \quad V = \begin{bmatrix} u \\ 0 \\ 0 \\ v \end{bmatrix},
\]
where \(g, v, \) and \(u\) are positive real numbers. Note that \(|\beta| \leq 1\) holds, because \(L^{*}L = I_{m} + i\tilde{J}\) is nonnegative definite. Then, (48) and (49) write
\[
u + v\lambda^2 - u^2 = 0, \quad vg^2 + u\lambda^2 - v^2g^2 = 0, \quad vg\lambda + u\lambda + uv\beta g = 0.
\]
The necessary and sufficient condition for \(\lambda\) and positive \(g\) to exist is, after some calculations,
\[
(u - 1)^{1/2} + (v - 1)^{1/2} - |\beta|(uv)^{1/2} = 0. \tag{54}
\]
Note that \(u\) and \(v\) are larger than or equal to 1, because \(V \geq JS - 1 = I_{m}\).
Substitution of \(u = z + x\) and \(v = z - x\) into (54), after some calculations, leads to
\[
|\beta| = |\beta|(z + x - 1)^{1/2}(z - x - 1)^{1/2} \pm (1 - \beta^2)^{1/2}((z + x - 1)^{1/2} + (z - x - 1)^{1/2}) \tag{55}
\]
Fig.1 shows that the lower sign in the equation (55),
\[
|\beta| = |\beta|(z + x - 1)^{1/2}(z - x - 1)^{1/2} + (1 - \beta^2)^{1/2}((z + x - 1)^{1/2} + (z - x - 1)^{1/2}) \tag{56}
\]
gives a part of \(b d\tilde{\hat{V}}\). In (56), \(x\) takes value ranging from \(-\beta^2/(1 - \beta^2)\) to \(\beta^2/(1 - \beta^2)\) if \(|\beta|\) is smaller than 1. When \(|\beta| = 1\), \(x\) varies from \(-\infty\) to \(\infty\). This restriction on the range of \(x\) comes from the positivity of \(z - x - 1\) and \(z + x - 1\).
When the weight matrix \(G\) is
\[
G = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{or} \quad G = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \tag{57}
\]
we must treat the case of \(|\beta| = 1\) and the case of \(|\beta| < 1\) separately. If \(|\beta| = 1\), there exists no \((2m + 1) \times m\) complex matrix \(X\) which satisfies \(V = \text{Re}X^{*}X\), (47), (33), and (36). On the other hand, if \(|\beta| < 1\), such
complex matrix $X$ always exists and $V = \text{Re}X^*X$ is given by, in terms of $(x, y, z)$,

\[ z = -x + 1, \quad x \leq -\frac{\beta^2}{1-\beta^2} \quad \text{or} \quad z = x + 1, \quad x \geq \frac{\beta^2}{1-\beta^2} \quad (58) \]

With the help of (56) and (58), $V$ is depicted as Fig. 2. The intersection of $z$-axis and $bdV$ gives

\[ \text{CR}(J^S) = \min\{J^SV | V \in \mathcal{V}\} = \frac{4}{1 + (1-|\beta|^2)^{1/2}}, \quad (59) \]

where the equality holds in any coordinate of the model $\mathcal{M}$. Simple calculation leads to following theorem.

**Theorem 26:** If a model $\mathcal{M}$ has larger value of $|\beta|$ at $\theta$ than another model $\mathcal{N}$ has at $\theta'$, the $\mathcal{V}_\theta(\mathcal{M})$ is a subset of $\mathcal{V}_{\theta'}(\mathcal{N})$.

By virtue of this theorem, $|\beta|$ can be seen as a measure of ‘uncertainty’ between the two parameters. Two extreme cases are worthy of special attention; When $|\beta| = 0$, the model $\mathcal{M}$ is locally quasi-classical at $\theta$ and $V$ is largest. On the other hand, if $|\beta| = 1$, $V$ is smallest and uncertainty
between $\theta^1$ and $\theta^2$ is maximum. In the latter case, we say that the model is coherent at $\theta$.

\begin{align*}
(b) \quad & \begin{array}{c}
\includegraphics[width=0.5\textwidth]{fig2}\end{array} \\
(a) \quad & \begin{array}{c}
\includegraphics[width=0.5\textwidth]{fig2}\end{array} \\
(c) \quad & \begin{array}{c}
\includegraphics[width=0.5\textwidth]{fig2}\end{array}
\end{align*}

Fig. 2. (a) $|\beta|=0$; (b) $0<|\beta|<1$; (c) $|\beta|=1$.

**Example 27:** We define *generalized spin coherent model* [2] by

\[
\begin{align*}
\mathcal{M}_{j,m} &= \pi(\mathcal{N}) \\
\mathcal{N}_{j,m} &= \{ |\phi(\theta)\rangle | |\phi(\theta)\rangle = \exp i\theta^1 (\sin \theta^2 S_x - \cos \theta^2 S_y)\rangle_{j,m}, \\
& \quad 0 \leq \theta^1 < \pi, \ 0 \leq \theta^2 < 2\pi \}
\end{align*}
\]

(60)

where $S_x, S_y, S_z$ are spin operators, and $|j,m\rangle$ is defined by,

\[
S_z |j,m\rangle = \hbar m |j,m\rangle, \quad (S_x^2 + S_y^2 + S_z^2)|j,m\rangle = \hbar^2 j(j+1)|j,m\rangle.
\]

$j$ takes value of half integers, and $m$ is a half integer such that $-j \leq m \leq j$. 
Then after tedious calculations, we obtain
\[ M \left( h |_{j,m} \partial_1 \right) = 2i(\sin^2 S_x - \cos^2 S_y)|j, m\rangle \]
\[ M \left( h |_{j,m} \partial_2 \right) = 2i(-\sin \theta_1 \cos S_x + \sin \theta^2 S_y + (\cos \theta^1 - 1)(S_z - m\hbar))|j, m\rangle, \]
and
\[ J^S = 2\hbar^2 (j^2 + j - m^2) \begin{bmatrix} 1 & 0 \\ 0 & \sin^2 \theta^1 \end{bmatrix}, \quad \tilde{J} = \begin{bmatrix} 0 & 2m\hbar^2 \sin \theta^1 \\ -2m\hbar^2 \sin \theta^1 & 0 \end{bmatrix}, \]
\[ \beta_{j,m} = \frac{m}{j^2 + j - m^2}. \]

If \( m = \alpha j \), where \( \alpha < 1 \) is a constant, \( \beta_{j,m} \) tends to zero as \( j \to \infty \), and the model \( M_{j,m} \) becomes locally quasi-classical. However, if \( m = j \), the model \( M_{j,m} \) is coherent for any \( j \).

### 5.7. Multiplication of the imaginary unit

As is shown in the previous subsection, in the 2-parameter model, \( |\beta| \), a good index of ‘uncertainty’ between two parameters, or a measure of how distinct the model is from the classical model. It can be easily shown that, whatever coordinate of the model \( M \) is chosen, \( \pm i\beta \) are the eigenvalues of the matrix \( J^S \tilde{J} \), which is deeply related to the complex structure of the model. Actually, that matrix stands for the linear map \( D \) from \( T_\rho(M) \) onto \( T_\rho(M) \) defined as in the followings; First, we multiply the imaginary unit \( i \) to \( |l_X\rangle = M(h(|l_X\rangle)) \) and \( M^{-1} \) and \( \pi_\ast \) are applied successively to \( i|l_X\rangle \). Since \( \pi_\ast(M^{-1}(i|l_X\rangle)) \) is not a member of \( T_\rho(M) \) generally, we project \( \pi_\ast(M^{-1}(i|l_X\rangle)) \in T_\rho(P_1) \) to \( T_\rho(M) \) with respect to the metric \( \langle \ast | \ast \rangle_\rho \), and we obtain \( DX \in T_\rho(M) \).

multiplication of \( i \)
\[
\begin{array}{cccc}
|l_X\rangle \in \text{span}_{\mathbb{R}}L & \longrightarrow & i|l_X\rangle \in \text{span}_{\mathbb{R}}\{L, iL\} \\
\uparrow & & \downarrow M^{-1}, \pi_\ast & \\
h|\phi\rangle, M & \pi_\ast(M^{-1}(i|l_X\rangle)) \in T_\rho(P_1) & \text{project} \downarrow \text{w.r.t.} \langle \ast | \ast \rangle_\rho \\
X \in T_\rho(M) & \longrightarrow & DX \in T_\rho(M) & D
\end{array}
\]

The following theorems are straightforward consequences of the above discussion.
Theorem 28: The absolute value of an eigenvalues of $D$, or equivalently, of $J^{S-1}\tilde{J}$, is smaller than or equal to 1.

5.8. The coherent model

As for the model with arbitrary number of parameters, the model is said to be coherent at $\theta$ iff all of the eigenvalues of $(J^S)^{-1}\tilde{J}$ are $\pm i$. When the number of parameters is 2, this definition of coherency reduces to $|\beta|=1$.

It should be noted that the eigenvalues of $J^{S-1}\tilde{J}$ are, whether the model is coherent or not, of the form $\pm i\beta_j$ or 0. Therefore, the number of parameters of the coherent model is even.

In this subsection, we determine the attainable CR type bound of the coherent model with arbitrary numbers of parameters. The coherent model is worthy of attention firstly because the coherent model is ‘the maximal uncertainty’ model, secondly because there are many physically important coherent models.

Because $J^{S-1}\tilde{J}$ is a representation of $D$, or of multiplication of the imaginary unit $i$, the following theorem.

Theorem 29: The model $\mathcal{M}$ is coherent at $\theta$ iff

$$ (J^{S-1}\tilde{J})^2 = -I_m $$

holds true.

Theorem 30: The model $\mathcal{M}$ is coherent at $\theta$ iff $\text{span}_R\{iL\}$ is identical to $\text{span}_R L$, or equivalently, iff $\text{span}_R\{L, iL\}$ is identical to $\text{span}_R L$.

This theorem leads to the following theorem.

Theorem 31: The model $\mathcal{M}$ is coherent at $\theta$ iff the dimension of $\text{span}_C L$ is $m/2$.

Proof: First, we assume that

$$ \dim_C \text{span}_C L = m/2. $$

Because $\text{span}_R L$ is a $m$-dimensional subspace of $\text{span}_R\{L, iL\}$ whose dimension is smaller than or equal to $m$ because of (62), we have $\text{span}_R\{L, iL\} = \text{span}_R L$, or coherency of the model at $\theta$.

Conversely, let us assume that the model is coherent at $\theta$. If we take an orthonormal basis $\{e_j | j = 1, ..., m\}$ of $T_\rho(\mathcal{M})$ such that $e_{j+m/2} = De_i (i =...$
1, 2, ..., \(m/2\), then \(M(h(e_{j+m/2})) = iM(h(e_j))(i = 1, 2, ..., m/2)\) holds true, and any element \(|u\rangle\) of \(\text{span}_{\mathbb{R}}L = \text{span}_{\mathbb{R}}\{L, iL\}\) writes

\[
|u\rangle = \sum_{j=1}^{m} a_j M(h(e_j)) = \sum_{j=1}^{m/2} (a_j + ia_{j+(m/2)}) M(h(e_j)),
\]

implying that the dimension of \(\text{span}_{\mathbb{C}}L\) is \(m/2\).

In 1996, Fujiwara and Nagaoka [9] determined the attainable CR type bound of the two parameter coherent model. In the following, more generally, we calculate the bound of the coherent model with arbitrary number of parameters.

**Lemma 32:** In the case of the coherent model, \(\text{Re} L^* X = I_m\) or its equivalence \(\text{Re} L^* (X - L J^{S-1}) = 0\), implies

\[
L^* X = I_m + i\tilde{J} J^{S-1}.
\]

**Proof:**

\[
\text{Im} L^* (X - L J^{S-1}) = -\text{Re} iL^* (X - L J^{S-1}) = 0.
\]

Here, \(\text{span}_{\mathbb{R}}L = \text{span}_{\mathbb{R}}iL\) is used to deduce the last equality. (63) and \(\text{Re} L^* (X - L J^{S-1}) = 0\) implies

\[
L^* X = L^* L J^{S-1} = I_m + i\tilde{J} J^{S-1}.
\]

Multiplication of \(L^*\) to the both sides of (46), together with the lemma presented above, yields

\[
(I_m + i\tilde{J} J^{S-1})(G - i\Lambda) = (J^S + i\tilde{J}) VG.
\]

By virtue of (61), both of the real part and the imaginary part of (64) give the same equation,

\[
G + \tilde{J} J^{S-1} \Lambda = J^S VG,
\]

or

\[
G^{1/2} VG^{1/2} - G^{1/2} J^{S-1} G^{1/2} = \left( G^{1/2} J^{S-1} \tilde{J} J^{S-1} G^{1/2} \right) \left( G^{-1/2} \Lambda G^{-1/2} \right).
\]
Therefore, letting $a_i$ and $b_i$ denote the eigenvalues of $G^{1/2}J S^{-1} \ddot{J}J S^{-1}G^{1/2}$ and $G^{-1/2} \Lambda G^{-1/2}$ respectively, we have
\[
\left[ (G^{1/2}J S^{-1} \ddot{J}J S^{-1}G^{1/2}), (G^{-1/2} \Lambda G^{-1/2}) \right] = 0
\]

\[
\operatorname{Tr} \left\{ (G^{1/2}J S^{-1} \ddot{J}J S^{-1}G^{1/2}) (G^{-1/2} \Lambda G^{-1/2}) \right\} = \sum_i |a_i||b_i|,
\]
because $|a_i| |b_i|$ implies that $(G^{1/2}J S^{-1} \ddot{J}J S^{-1}G^{1/2}) \times (G^{-1/2} \Lambda G^{-1/2})$ is positive Hermitian.

On the other hand, (46) or its equivalence,
\[
XG^{1/2}(I_m - iG^{-1/2} \Lambda G^{-1/2}) = LVG^{1/2},
\]
implies $|b_i| = 1 (i = 1, ..., m)$, because the rank of $I_m - iG^{-1/2} \Lambda G^{-1/2}$ is shown to be $m/2$ from (66), and the rank of matrices $L$, $X$, and $V$.

\[
\operatorname{rank}_{\mathbb{C}} X = \operatorname{rank}_{\mathbb{R}} X = m,
\]
hold true, where the last equation is valid by virtue of $\operatorname{Im} X^* X = 0$.

After all, letting $\operatorname{Tr} \abs A$ denote the sum of the absolute values of the eigenvalues of $A$, we have the following theorem.

**Theorem 33:**
\[
\operatorname{CR}(G) = \operatorname{Tr} G J S^{-1} + \operatorname{Tr} \abs G J S^{-1} \ddot{J}J S^{-1},
\]
where letting $|A| = (AA^*)^{1/2}$, the covariance matrix $V$ such that
\[
V = J S^{-1} + G^{-1/2} \left| G^{1/2}J S^{-1} \ddot{J}J S^{-1}G^{1/2} \right| G^{-1/2}.
\]
attain the minimum.

To check the coherency of the model, the following theorem, which is induced from theorem 28, is useful.

**Theorem 34:** the model is coherent at $\theta$ iff
\[
| \det J S | = | \det \ddot{J} |.
\]

**Example 35:** (squeezed state model) Squeezed state model, which has four parameters, is defined by
\[
\mathcal{M} = \{ \rho(z, \xi) \mid \rho(z, \xi) = \pi(|z, \xi \rangle), z, \xi \in \mathbb{C} \},
\]
where
\[
|z, \xi \rangle = D(z)S(\xi)|0\rangle, \quad D(z) = \exp(za^\dagger - \bar{z} a), \quad S(\xi) = \exp \left\{ \frac{1}{2} (\xi a^\dagger a - \bar{\xi} a^2) \right\}.
\]
Here, the operator $a$ is defined as $a = (Q + iP)/\sqrt{2\hbar}$ where $P$ and $Q$ satisfy the canonical commutation relation $[P, Q] = -i\hbar$. Letting $z = \sqrt{\frac{2}{\hbar}}(\theta^1 + i\theta^2)$, and $\xi = \theta^3 e^{-2i\theta^4}$, we have

$$M \left( \hbar \partial_{\theta^1} \right) |z, \xi\rangle = \frac{2i}{\hbar} (P + \theta^2) |z, \xi\rangle,$$

$$M \left( \hbar \partial_{\theta^3} \right) |z, \xi\rangle = -i(e^{2i\theta^4} a^2 - e^{-2i\theta^4} a^\dagger^2) |z, \xi\rangle,$$

$$M \left( \hbar \partial_{\theta^4} \right) |z, \xi\rangle = i4(\sinh^2 \theta^3)(a^\dagger a + \frac{1}{2}) |z, \xi\rangle,$$

and

$$J^S = \frac{2}{\hbar} \begin{bmatrix}
\cosh 2\theta^3 - \sinh 2\theta^3 \cos 2\theta^4 & \sinh 2\theta^3 \sin 2\theta^4 & 0 & 0 \\
\sinh 2\theta^3 \sin 2\theta^4 & \cosh 2\theta^3 + \sinh 2\theta^3 \cos 2\theta^4 & 0 & 0 \\
0 & 0 & \hbar & 0 \\
0 & 0 & 0 & \hbar \sinh 2\theta^3
\end{bmatrix},$$

$$\tilde{J} = \frac{2}{\hbar} \begin{bmatrix}
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{\hbar}{2} \sinh 2\theta^3 \\
0 & 0 & \frac{\hbar}{2} \sinh 2\theta^3 & 0
\end{bmatrix}.$$

Coherency of this model is easily checked by theorem 34,

$$|\det J^S| = |\det \tilde{J}| = \frac{4}{\hbar^2} \sinh^2 2\theta^3.$$

**Example 36:** (spin coherent model) As is pointed out by Fujiwara [9], spin coherent model $\mathcal{M}_{s,s}$, where $\mathcal{M}_{s,m}$ is defined by (60), is coherent.

**Example 37:** (total space model) The total space model is the space of all the pure state $P_1$ in finite dimensional Hilbert space $\mathcal{H}$. By virtue of theorem 30, the coherency of the model is proved by checking that $\text{span}_{\mathcal{L}}$ is invariant by the multiplication of the imaginary unit $i$. Let $|l\rangle$ be a horizontal lift of a tangent vector at $|\phi\rangle$. Then, $i|l\rangle$ is also a horizontal lift of another tangent vector at $|\phi\rangle$, because $|\phi\rangle + i|l\rangle dt$ is a member of $\mathcal{H}$.

### 5.9. Informationally exclusive, and independent parameters

In a $m$-parameter model $\mathcal{M}$, we say parameter $\theta^i$ and $\theta^j$ are informationally independent at $\theta_0$, iff

$$\text{Re} \langle l_i | l_j \rangle_{\theta = \theta_0} = \text{Im} \langle l_i | l_j \rangle_{\theta = \theta_0} = 0,$$
because if the equation holds true, in the estimation of the parameters $\theta^1, \theta^2$ of the 2-parameter submodel $\mathcal{M}(1, 2|\theta_0)$ of $\mathcal{M}$, where

$$\mathcal{M}(1, 2|\theta_0) \equiv \{ \rho(\theta) \mid \theta = (\theta^1, \theta^2, \theta^3_0, \ldots, \theta^m_0), (\theta^1, \theta^2) \in \mathbb{R}^2 \}, \quad (67)$$

both of the parameters can be estimated up to the accuracy which is achieved in the estimation of the parameter of the 1-parameter submodels $\mathcal{M}(1|\theta_0)$ and $\mathcal{M}(2|\theta_0)$, where

$$\mathcal{M}(1|\theta_0) \equiv \{ \rho(\theta) \mid \theta = (\theta^1, \theta^2_0, \ldots, \theta^m_0), \theta^1 \in \mathbb{R} \},$$

$$\mathcal{M}(2|\theta_0) \equiv \{ \rho(\theta) \mid \theta = (\theta^1_0, \theta^2, \ldots, \theta^m_0), \theta^2 \in \mathbb{R} \}.$$

On the other hand, iff

$$\text{Re} \langle l_1|l_2 \rangle|_{\theta=\theta_0} = 0, \quad (68)$$

and $\mathcal{M}(1, 2|\theta_0)$ is coherent, or equivalently,

$$\text{Im} \langle l_1|l_2 \rangle|_{\theta=\theta_0} = \left( \langle l_1|l_1 \rangle \langle l_2|l_2 \rangle \right)^{1/2}|_{\theta=\theta_0}, \quad (70)$$

hold true, we say the parameters are \textit{informationally exclusive} at $\theta_0$.

Fujiwara and Nagaoka [9] showed that in the coherent model with two orthogonal parameters, the attainable CR type bound is achieved by applying the best measurement for each parameter alternatively to the system. This fact implies that if two parameters are informationally exclusive, the one of them do not contain any information about the other. In fact, we have the following theorem.

**Theorem 38:** If two parameters $\theta^1$ and $\theta^2$ are informationally exclusive, any unbiased measurement $M$ in $\mathcal{M}(1, 2|\theta^0)$ which estimates $\theta^1$ as accurately as possible, i.e.,

$$\int (\hat{\theta}^1 - \theta^1)^2 \text{Tr} \rho(\theta_0) M(d\hat{\theta}) = [J^{S-1}]^{11} \quad (69)$$

can extract no information about $\theta^2$ from the system, i.e.,

$$\forall B \subset \mathbb{R}^2 \quad \text{Tr} \left( M(B) \frac{\partial \rho}{\partial \theta^2} \bigg|_{\theta=\theta_0} \right) = \text{Re} \langle \phi|M(B)|l_2 \rangle = 0, \quad (70)$$

and vice versa.

**Proof:** We prove the theorem only for the measurements which writes

$$M(B) = \int_B |\hat{\theta}\rangle \langle \hat{\theta}| \mu(d\hat{\theta}).$$
The proof for general case will be discussed elsewhere. If (69) holds true, as in the proof of lemma 4 (see p. 88 [12]),

\[ \langle \hat{\theta} | \left\{ (\hat{\theta}^1 - \theta^1) | \phi \right\} - \frac{1}{\langle l_1 | l_1 \rangle} | l_1 \rangle = 0 \]  

(71)
must hold. On the other hand, because of coherency of \( M(1, 2 | \theta^0) \) and (68), for some real number \( a \), we have

\[ | l_2 \rangle = ia | l_1 \rangle, \]

by use of which it is shown that (70) is equivalent to

\[ \text{Im}( \langle \phi | \hat{\theta} \langle \hat{\theta} | l_1 \rangle ) = 0. \]

This equation is obviously true if (71) is true, and we have the theorem. ■

5.10. Direct sum of models

For the submodels

\[ M_1 \equiv M(1, 2, ..., m_1 | \theta^0), \quad M_2 \equiv M(m_1, m_1 + 1, ..., m | \theta^0) \]

of \( M \), where \( M(1, 2, ..., m_1 | \theta^0) \) and \( M(m_1, m_1 + 1, ..., m | \theta^0) \) are defined almost in the same way as the definition (67) \( M(1, 2 | \theta^0) \), we write

\[ M|_{\theta^0} = M_1 \oplus M_2|_{\theta^0}, \]

and say that \( M \) is sum of \( M \) and \( M \) at \( \theta^0 \). \( m - m_1 \) is denoted by \( m_2 \).

Lemma 39: If any parameter of \( M_1 \) is informationally independent of any parameter of \( M_2 \) at \( \theta^0 \), and the weight matrix \( G \) writes

\[ G = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}, \]

then

\[ \text{CR}(G, \theta^0, M) = \text{CR}(G_1, \theta^0, M_1) + \text{CR}(G_2, \theta^0, M_2). \]

When the assumption of the lemma is satisfied, \( M_1 \) and \( M_2 \) are said to be informationally independent at \( \theta^0 \).

Proof: Let \( \hat{\theta}_1(\omega) \) and \( \hat{\theta}_2(\omega) \) be the vector whose components are the estimates of \( \theta^1, \theta^2, ..., \theta^{m_1} \), and \( \theta^{m_1+1}, \theta^{m_1+2}, ..., \theta^m \),

\[ \hat{\theta}_1(\omega) = \left( \hat{\theta}^1(\omega), \hat{\theta}^2(\omega), ..., \hat{\theta}^{m_1}(\omega) \right), \quad \hat{\theta}_2(\omega) = \left( \hat{\theta}^{m_1+1}(\omega), \hat{\theta}^{m_1+2}(\omega), ..., \hat{\theta}^m(\omega) \right). \]
Then, if \((M, \hat{\theta}, \Omega)\) is locally unbiased, \((M, \hat{\theta}_i, \Omega)\) is locally unbiased. Therefore, we have,

\[
\{ M \mid (M, \exists \hat{\theta}_i, \Omega) \text{ is locally unbiased} \} \supset \{ M \mid (M, \exists \hat{\theta}, \Omega) \text{ is locally unbiased} \}
\]

which yields,

\[
\min \left\{ \text{Tr} \, G V [\hat{\theta} | M] \mid (M, \hat{\theta}, \Omega) \text{ is locally unbiased} \right\} = \min \left\{ \text{Tr} \, G_1 V [\hat{\theta}_1 | M] \mid (M, \hat{\theta}, \Omega) \text{ is locally unbiased} \right\} + \min \left\{ \text{Tr} \, G_2 V [\hat{\theta}_2 | M] \mid (M, \hat{\theta}, \Omega) \text{ is locally unbiased} \right\},
\]

or its equivalence,

\[
\text{CR}(G, \mathcal{M}) \geq \text{CR}(G_1, \mathcal{M}_1) + \text{CR}(G_2, \mathcal{M}_2). \tag{72}
\]

Because \(\mathcal{M}_1\) and \(\mathcal{M}_2\) are informationally independent, \(L\) for \(\mathcal{M}\) writes

\[
L = \begin{bmatrix} L_1 & 0 \\ 0 & L_2 \end{bmatrix},
\]

in the appropriate coordinate, where \(L_1 = [|l_1\rangle, |l_2\rangle, ..., |l_{m_1}\rangle]\), and \(L_2\) is in the same manner. In that coordinate, \(X\) writes

\[
X = \begin{bmatrix} X_1 & X_{12} \\ X_{21} & X_2 \end{bmatrix}.
\]

Therefore, if

\[
\text{Re} X_i^* L_i = I_{m_i}, \quad \text{Im} X_i^* X_i = 0 (i = 1, 2)
\]

holds true, the measurements corresponding to \(X\) is locally unbiased, and

\[
\text{CR}(G, \mathcal{M}) = \min \left\{ \text{Tr} \, G X^* X \mid \text{Re} X^* L = I_m, \text{Im} X^* X = 0 \right\} \leq \min \left\{ \sum_{i=1}^{2} \text{Tr} \, G_i X_i^* X_i \mid \text{Re} X_i^* L_i = I_{m_i}, \text{Im} X_i^* X_i = 0, (i = 1, 2) \right\} = \text{CR}(G_1, \mathcal{M}_1) + \text{CR}(G_2, \mathcal{M}_2),
\]

which, mixed with (72) leads to the lemma.
6. Berry’s Phase in Quantum Estimation Theory

6.1. Berry’s phase

In this subsection, we review the geometrical theory of Berry’s phase.

Berry’s phase was discovered by M.V. Berry in 1984 [4], and confirmed by many experimental facts [22]. In 1987, Aharonov and Anandan [1] pointed out that Berry’s phase is naturally interpreted as a curvature in the fiber bundle over $\mathcal{P}_1$. Actually, Berry’s phase is nothing but the Uhlmann’s curvature restricted to $\mathcal{P}_1$ [24, 26].

Uhlmann’s RPF in the space $\mathcal{P}_1$ takes value in the set of unimodular complex number. On the other hand, Berry’s phase takes value in real numbers. They are related as

\[ \text{Berry’s phase} = -i \ln(\text{Uhlmann’s RPF}). \]

The Berry’s phase for the infinitesimal loop (16) is calculated up to the second order of $d\theta$ as

\[ \frac{1}{2i} \langle \phi(0) | F_{ij} | \phi(0) \rangle d\theta^i d\theta^j + o(d\theta)^2 \]

Because Berry’s phase is independent of the choice of SLD, we can take $L_i^S$ to be $2\partial_i \rho$. Then, the phase is equal to

\[ \frac{1}{2} \tilde{J}_{ij} d\theta^i d\theta^j + o(d\theta)^2, \]

where $\tilde{J}_{ij}$ is equal to $\frac{1}{2} \text{Im} \langle l_i | l_j \rangle$.

Mathematically,

\[ \sum_{i,j} \tilde{J}_{ij} d\theta^i d\theta^j \quad (73) \]

corresponds to the curvature form.

6.2. Berry’s phase in quantum estimation theory

It must be noted that the curvature form is deeply related to the multiplication of the imaginary unit $D$. Actually, The curvature form (73) is identical to a map from $T_\theta(\mathcal{M}) \times T_\theta(\mathcal{M})$ to $\mathbb{R}$ such that

\[ \sum_{ij} \tilde{J}_{ij} d\theta^i d\theta^j : (\partial_i, \partial_j) \longrightarrow \left\langle \partial_i, \frac{1}{2} D \partial_j \right\rangle. \]

Hence, the eigenvalues of $D$ can be interpreted in terms of Berry’s phase. Concretely speaking, taking the coordinate system which is orthonormal at
θ in terms of the metric $\langle *, * \rangle_\theta$, they are the half of the Berry’s phase obtained when the state goes around the infinitesimal loop (16). Especially, when the model is only with two parameters, the eigenvalues of $\mathbf{D}$ are the half of the Berry’s phase per unit area, where unit of the area is naturally induced from the metric $\langle *, * \rangle_\theta$.

Therefore, we can roughly say that the more the Berry’s phase for the loop (16), the harder it is to estimate $\theta^i$ and $\theta^j$ simultaneously. Namely, $\theta^i$ and $\theta^j$ are informationally independent iff the Berry’s phase for the loop (16) vanishes and $\langle \partial_i, \partial_j \rangle_\theta = 0$ holds. On the other hand, iff the Berry’s phase for the loop (16) is maximal and $\langle \partial_i, \partial_j \rangle_\theta = 0$ holds, i.e., $\beta = 1$, the two parameters are informationally exclusive.

This discussion is parallel to that in the subsection 4.2, which was about relations between Uhlmann’s parallelism and the noncommutative nature of the quantum estimation theory of the faithful model. Because Berry’s phase is nothing but the restriction of Uhlmann’s RPF to the pure state model, this parallelism is natural.

What about the models with arbitrary number of parameters? By virtue of theorem 18, if Berry’s phase for any closed loop vanishes, the model is locally quasi-classical. For general pure state models, we have the following theorem.

**Theorem 40:** For any pure state model,

\[
\text{CR}(J^S) = \text{Tr} \left\{ \text{Re} (I_m + ij^{S-1/2} \tilde{J} J^{S-1/2})^{1/2} \right\}^{-2} = \sum_{\alpha \in \{ \text{eigenvalues of } \mathbf{D} \}} \frac{2}{1 + (1 - |\alpha|^2)^{1/2}}.
\]

The estimation theoretical significance of $\text{CR}(J^S)$ is hard to verify. However, this value remains invariant under any transform of the coordinate in the model $\mathcal{M}$, and can be a good index of distance between $\mathcal{V}$ and $J^{S^{-1}}$.

**Proof:** Because $\text{CR}(J^S)$ is invariant by any affine coordinate transform in the model $\mathcal{M}$, we choose a coordinate in which $J^S$ writes $I_m$ and $\tilde{J}$ writes
Then, the model $\mathcal{M}$ is decomposed into the direct sum of the submodels one or two parameter $\mathcal{M}_\kappa$,

$$\mathcal{M} = \bigoplus_\kappa \mathcal{M}_\kappa,$$

where any two submodels $\mathcal{M}_\kappa$ and $\mathcal{M}_{\kappa'}$ are informationally independent, and $\tilde{J}$ of a two parameter submodel $\mathcal{M}_\kappa$ is

$$\tilde{J} = \begin{bmatrix}
0 & -\beta_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\beta_1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 0 & 0 & 0 & -\beta_l & \cdots & 0 \\
0 & 0 & \cdots & 0 & \beta_l & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\end{bmatrix}.$$

Because the weight matrix $J^S = I_m$ writes in the form of direct some of the weight matrix on the model $\mathcal{M}_\kappa$, by virtue of lemma 39 and the equation (59), we have the theorem.

\section*{6.3. Berry’s phase in the global theory of quantum estimation}

In this subsection, we present a geometrical sufficient condition for the pure state model $\mathcal{M}$ to be quasi-classical in the sense of subsection 3.6.

For simplicity, we say that the manifold $\mathcal{N}$ in $\mathcal{P}_1$ is a horizontal lift of the model $\mathcal{M}$ if

$$\pi(\mathcal{N}) = \mathcal{M}, \quad \forall |\phi(\theta)\rangle \in \mathcal{N}, \quad \frac{1}{2} \frac{\partial}{\partial \theta^i} |\phi(\theta)\rangle \in \mathcal{L}S_{|\phi(\theta)\rangle}.$$

The horizontal lift $\mathcal{N}$ exists iff $\mathcal{M}$ is quasi-classical.

\textbf{Theorem 41:} If $\mathcal{H}$ is finite dimensional and the model $\mathcal{M}$ is parallel, that model is globally classical in the sense of subsection 3.6.
Proof: Because the model is parallel, Berry’s phase for any closed loop also vanishes, i.e., there exists a horizontal lift $\mathcal{N}$ of the model $\mathcal{M}$.

For the model $\mathcal{M}$ to be parallel, any members $|\phi(\theta)\rangle$, $|\phi(\theta')\rangle$ of the horizontal lift $\mathcal{N}$ must satisfy

$$
\langle \phi(\theta) | \phi(\theta') \rangle \in \mathbb{R},
$$

which implies that $\mathcal{N}$ is a subset of the real separable Hilbert space $\mathcal{R}$ spanned by the appropriately chosen basis $\{|e_j\rangle | j = 1, 2, ..., d\}$ in $\mathcal{H}$,

$$
\mathcal{N} \subset \mathcal{R} = \left\{ |\phi\rangle \bigg| |\phi\rangle = \sum_{j} a^j |e_j\rangle \right\}.
$$

(74)

We claim that the optimal estimator is constructed by use of the measurement

$$
M(B) = \int_{B} O|e_1\rangle \langle e_1| O^T \mu(dO),
$$

where $O$ is an orthogonal transform in $\mathcal{R}$, and $\mu(\ast)$ is the invariant measure in the space of all the orthogonal transform in $\mathcal{R}$.

To prove this claim, in the following, we calculate the classical Fisher information matrix of the probability distribution family

$$
\left\{ p(O|\theta) \bigg| \int_{B} p(O|\theta) \mu(dO) = \int_{B} \langle \phi(\theta)|M(dO)|\phi(\theta)\rangle \big| \theta \in \Theta \right\}.
$$

From the definition, the $(i, j)$ th component of the classical Fisher information matrix is

$$
\int_{\Omega} \mu(dO) \frac{\partial_i p(O|\theta) \partial_j p(O|\theta)}{p(O|\theta)} = \int_{\Omega} \mu(dO) \frac{4\langle \partial_i \phi(\theta)|O|e_1\rangle \langle e_1|O^T \phi(\theta)\rangle \langle \partial_j \phi(\theta)|O|e_1\rangle \langle e_1|O^T \phi(\theta)\rangle}{\langle \phi(\theta)|O|e_1\rangle \langle e_1|O^T \phi(\theta)\rangle} = \int_{\Omega} \mu(dO) 4\langle \partial_i \phi(\theta)|O|e_1\rangle \langle \partial_j \phi(\theta)|O|e_1\rangle
$$

$$
= 4\langle \partial_i \phi(\theta)\rangle \int_{\Omega} \mu(dO) \langle e_1|O^T \partial_j \phi(\theta)\rangle \langle e_1|O \partial_j \phi(\theta)\rangle = 4\langle \partial_i \phi(\theta)|\partial_j \phi(\theta)\rangle = \langle l_i|l_j\rangle,
$$

which is nothing but the $(i, j)$ th component of the SLD Fisher information matrix, because $\langle l_i|l_j\rangle = 4\langle \partial_i \phi(\theta)|\partial_j \phi(\theta)\rangle$ is real by virtue of (74).

We define the model is quasi-classical in the wider sense iff for any $\theta \in \Theta$, there is an open ball $B(\theta, \varepsilon)$ in $\mathbb{R}^m$ such that the model

$$
\mathcal{M}(\theta, \varepsilon) = \{ \rho(\theta) | \rho(\theta) \in \mathcal{M}, \theta \in B(\theta, \varepsilon) \subset \Theta \}.$$
is quasi-classical.

Theorem 42: If the model \( M \) is parallel, that model is quasi-classical in the wider sense.

Proof: Let us define \( N \) and \( R \) in the same way as in the proof of theorem 41. Then, the optimal measurement in \( M(\theta, \varepsilon) \) for some \( \varepsilon \) is obtained by use of commuting theorem.

The converse of the latter theorem is, however, not true, because the following counter-examples exist.

Example 43: We consider the position shifted model which is defined by

\[
M_x = \pi(N_x)
\]

\[
N_x = \{ |\phi(\theta)\rangle | \langle \phi(\theta)| = \text{const.} \times (x - \theta)^2 e^{-(x-\theta)^2 + ig(x-\theta)}, \theta \in \mathbb{R} \},
\]

where \( c \) is a normalizing constant, \( g \) the function such that

\[
g(x) = \begin{cases} 
0 \ (x \geq 0), \\
\alpha \ (x < 0).
\end{cases}
\]

Then, as easily checked, \( N_x \) is a horizontal lift of the model \( M_x \), and \( \langle \phi(\theta)|\phi(\theta')\rangle \) is not real unless \( \alpha = n\pi \ (n = 0, 1, \ldots) \). However, SLD CR bound is uniformly attained by the measurement obtained by the spectral decomposition \( E_x(dx) = |x\rangle\langle x|dx \) of the position operator, where \( |x_0\rangle = \delta(x - x_0) \), as is checked by comparing SLD Fisher information of the model \( M_x \) and the classical Fisher information of the probability distribution family

\[
\{ p(x|\theta) | p(x|\theta) = |\langle \phi(\theta)|x\rangle|^2, \theta \in \mathbb{R} \}.
\]

Note that \( |\phi(\theta)\rangle \) is an eigenstate of the Hamiltonian

\[
H = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{\hbar^2}{m} \left( 2x^2 + \frac{1}{x^2} \right),
\]

whose potential has two wells with infinite height of wall between them.

Example 44: Let \( \mathcal{H} \) be \( L^2([0, 2\pi], \mathbb{C}) \), and define a one parameter model \( M \) such that,

\[
M = \pi(\mathcal{N})
\]

\[
\mathcal{N} = \{ |\phi(\theta)\rangle | \langle \phi(\theta)| = \text{const.} \times (2 - \cos \omega) e^{i\alpha(f(\omega-\theta) + \theta)}, \ (0 \leq \omega, \theta < 2\pi) \},
\]

(75)
where $\alpha$ is a real number and $f$ the function defined by

$$f(\omega - \theta) = \begin{cases} 
\omega - \theta & (\omega - \theta \geq 0) \\
\omega + 2\pi - \theta & (\omega - \theta < 0)
\end{cases}.$$ 

Physically, (75) is an eigenstate of the Hamiltonian $H$ such that,

$$H = -\frac{\hbar^2}{2m} \left( \frac{d}{d\omega} - i\alpha \right)^2 + \frac{A - B \cos \omega}{2 - \cos \omega},$$

which characterize the dynamics of an electron confined to the one-dimensional ring which encircles magnetic flux $\Phi = 2\pi \alpha c/e$, where $m$ is the mass of the electron, $-e$ the charge of the electron, $c$ the velocity of light, and $A, B$ the appropriately chosen constant.

It is easily checked that $N$ is a horizontal lift of the model $\mathcal{M}$, and that the model $\mathcal{M}$ is not parallel unless $\alpha = n\pi$ ($n = 0, 1, \ldots$). However, consider the projection valued measure $E_\omega$ such that

$$E_\omega(d\omega) = |\omega\rangle\langle\omega|d\omega,$$

where $|\omega_0\rangle = \delta(\omega - \omega_0)$. Then, it is easily checked that the classical Fisher information of the probability distribution family

$$\{p(\omega|\theta) | p(\omega|\theta) = |\langle \phi(\theta)|\omega \rangle|^2, 0 \leq \omega, \theta < 2\pi \}$$

is equal to the SLD Fisher information of $\mathcal{M}$.

### 6.4. Antiunitary operators

The transformation $A$

$$|\tilde{a}\rangle = A|a\rangle, \quad |\tilde{b}\rangle = A|b\rangle$$

is said to be antiunitary iff

$$\langle \tilde{a}|\tilde{b}\rangle = \overline{\langle a|b \rangle}, \quad A(\alpha|a\rangle + \beta|b\rangle) = \overline{\alpha}A|a\rangle + \overline{\beta}A|b\rangle,$$

where $\overline{x}$ means the complex conjugate of $x$.

Fix an orthonormal basis $B = \{|i\rangle | i = 1, 2, \ldots, d\}$, and we can then define antiunitary operator $K_B$ which takes complex conjugate of any components in this basis,

$$K_B \left( \sum_i \alpha_i |i\rangle \right) = \sum_i \overline{\alpha_i} |i\rangle.$$

For different basis $B, B'$, we have

$$K_B = U K'_B U^*,$$
where $U$ is a unitary operator corresponding to the change of the basis.

Suppose that any member of the manifold $N = \{|\phi\rangle\}$ in $\tilde{H}$ is invariant by the antiunitary operator $A$, and let $|\tilde{\phi}\rangle = A|\phi\rangle$, $|\tilde{\phi}'\rangle = A|\phi'\rangle$. Then, we have

$$\langle \phi | \phi' \rangle = \langle \tilde{\phi}' | \tilde{\phi} \rangle = \langle \phi' | \phi \rangle \in \mathbb{R}.$$  
Conversely, if $\langle \phi | \phi' \rangle$ is real for any $|\phi\rangle, |\phi'\rangle \in N$, by Schmidt’s orthonormalization, we can obtain the basis $B$ such that $N$ is subset of the real span of $B$, which means any member of $N$ is invariant by the antiunitary operator $K_B$.

Therefore, the premise of the statement of theorems 41-42 is satisfied iff the horizontal lift of the model is invariant by some antiunitary operator.

6.5. **Time reversal symmetry**

As an example of the antiunitary operator, we discuss time reversal operator (see pp. 266–282 [20]). The time reversal operator $T$ is an antiunitary operator in $L^2(\mathbb{R}^3, \mathbb{C})$ which transforms the wave function $\psi(x) \in L^2(\mathbb{R}^3, \mathbb{C})$ as:

$$T\psi(x) = \overline{\psi(x)} = K_{\{x\}} \psi(x).$$

The term ‘time reversal’ came from the fact that if $\psi(x, t)$ is a solution of the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m}\nabla^2 + V\right) \psi,$$

then $\overline{\psi(x, -t)}$ is also its solution.

The operator $T$ is sometimes called motion reversal operator, since it transforms the momentum eigenstate $e^{i\mathbf{p} \cdot \mathbf{x}/\hbar}$ corresponding to eigenvalue $\mathbf{p}$ to the eigenstate $e^{-i\mathbf{p} \cdot \mathbf{x}/\hbar}$ corresponding to eigenvalue $-\mathbf{p}$.

Define the position shifted model by

$$M_x = \{ \rho(\theta) | \rho(\theta) = \pi(\psi(x - x_0)), x_0 \in \mathbb{R}^3 \},$$

and suppose that any member of the horizontal lift $N_x$ of the model $M_x$ has time reversal symmetry. Then, since time reversal operator $T$ is antiunitary, the model $M_x$ is quasi-classical in the wider sense. The spectral decomposition of the position operator gives optimal measurement.

Now, we discuss the generalization of time reversal operator. The antiunitary transform

$$T_\alpha : e^{i\mathbf{p} \cdot \mathbf{x}/\hbar} \rightarrow e^{i\alpha(\mathbf{p})} e^{-i\mathbf{p} \cdot \mathbf{x}/\hbar}$$
is also called motion reversal operator, or time reversal operator.

If any member $\psi(x - x_0)$ of the horizontal lift $N_x$ of the position shifted model $M_x$ is invariant by the time reversal operator $T_\alpha$, 

$$\int_{\mathbb{R}^3} \psi(x - x_0) \overline{\psi(x - x'_0)} \, dx \in \mathbb{R}$$  \hspace{1cm} (76)

holds true for any $x_0, x'_0$, which is equivalent to the premise of theorems 41-42.

Conversely, if (76) holds true, Fourier transform of (76) leads to 

$$|\Psi(p)|^2 = |\Psi(-p)|^2,$$

where

$$\Psi(p) = \frac{1}{\sqrt{2\pi}} \int \psi(x) e^{-i p \cdot x / \hbar} \, dx.$$  

Therefore, any member of $N_x$ is transformed to itself by the time reversal operator $T_\alpha$ such that

$$T_\alpha : e^{i p \cdot x / \hbar} \rightarrow e^{i(\beta(p) + \beta(-p))} e^{-i p \cdot x / \hbar},$$

where

$$e^{i \beta(p)} = \frac{\Psi(p)}{|\Psi(p)|}.$$

**Theorem 45:** (76) is equivalent to the existence of the time reversal operator which transforms any member of the horizontal lift $N_x$ to itself.

**References**


PART IV

Group Symmetric Approach to Pure States Model

Chap. 22:  S. Massar and S. Popescu
“Optimal extraction of information from finite quantum ensembles”  . 356

Chap. 23:  M. Hayashi
“Asymptotic estimation theory for a finite-dimensional
pure state model”  . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 365

Chap. 24:  D. Bruß, A. Ekert, and C. Macchiavello
“Optimal universal quantum cloning and state estimation”  . . . . . . . 379

Chap. 25:  A.S. Holevo
“Bounds for generalized uncertainty of the shift parameter”  . . . . . . . 386
CHAPTER 21

Introduction to Part IV

1. Optimal Performance in the Covariant Model

Group-theoretical symmetry is a fundamental property in quantum physics, and provides a very powerful method to research several topics in quantum theory. Especially, when a group acts on the quantum system, the irreducibility simplifies the problems very much. Quantum statistical inference has several important models with group-theoretical symmetry, although classically such models have less importance. Thus, it is a natural research direction to apply group-theoretical symmetry to quantum statistical inference.

The first approach based on symmetry was developed by Helstrom [IV-1], and he treated the estimation problem of one-parametric covariant pure states families. Following his result, Holevo [IV-2,IV-3] formulated a quantum covariant state family and a covariant measurement in a more general setting. If we adopt Bayesian criterion with an invariant prior and an invariant loss function in the estimation of any covariant model, the optimal estimator can be given as the optimal covariant estimators which is the minimax estimator. This statement is called the quantum Hunt-Stein theorem [IV-2,IV-3], whose details are reviewed in section 3 in Chap. 23. Using this method, he obtained minimum errors of several important group covariant models. However, he did not explicitly focus on the case of $n$-identical-copies.

After this work, Holevo [Chap. 24] focused on the $n$-identical-copies setting in a one-parametric shifted pure states model, and evaluated the asymptotic behavior of the estimation error. His result has been extended by Holevo [IV-4]. Massar and Popescu [Chap. 22] focused on the same setting in the full pure states model in the quantum two-level system. They proved that the optimal error equals $\frac{1}{n+2}$ in the covariant setting, whereas Holevo implicitly obtained the same result in section 10 of chapter IV in Holevo’s textbook [IV-3]. Hayashi [Chap. 23] treated the same problem with a more general covariant setting in a full pure states model in any finite-dimensional system. Here, we should remark that Massar and Popescu [Chap. 22] and
Holevo [Chap. 24] treated only the error $1 - |\langle \phi, \hat{\phi} \rangle|^2$ between the true state $|\phi\rangle$ and the estimated state $|\hat{\phi}\rangle$; however Hayashi [Chap. 23] treated an arbitrary monotone function of $1 - |\langle \phi, \hat{\phi} \rangle|^2$ as the error. Furthermore, Vidal et al. [IV-5] treated the optimal error of the mixed states model in the two-level system with the $n$-identical copies under any covariant Bayes prior. Bagan et al. [IV-6] treated the asymptotic behaviour of its average error.

2. Optimization among Realizable POVMs

These optimal measurements, however, are not necessarily realizable because they are usually collective measurements on the joint system of many particles. On the other hand, if a quantum measurement on the whole system consists of the repetition of the same quantum measurement of a single-particle system, it is realized relatively easily. Massar and Popescu [Chap. 22] proved that the optimal error cannot be realized by a separable measurement in the finite-copies case of their model. However, Hayashi [Chap. 23] indicated that the optimal error can be asymptotically attained to the first order by the repetition of the optimal measurement of a single copy without the adaptive method. He also showed that there is a small gap concerning large deviation evaluation between these two settings.

In addition, it is easily verified that these covariant models are coherent models in the sense of [Chaps. 18,19]. Therefore, the optimal coefficients with the first asymptotic order of these models can be calculated by Fujiwara [Chap. 18], and Fujiwara and Nagaoka [Chap. 19]. Also, from the discussion in chapter 16, we can check that the optimal coefficients can be attained by an adaptive measurement.

3. Application to Other Problems

The group representation theoretical method is useful for analyzing the approximate cloning of a pure state with $N$ input qubits and $M$ output qubits [IV-7]. In particular, Bruß, Ekert, and Macchiavello [Chap. 25] focused on the relation between quantum estimation and quantum approximate cloning, and obtained the optimal performance of quantum approximate cloning in this case. This is a good application of quantum estimation, and indicates the importance of quantum estimation for the foundation of quantum information. This result has been extended to the $d$-dimensional case by Werner [IV-8]. These results indicate that quantum estimation can be regarded as the quantum cloning with infinite output qubits.
Another application of this method is the estimation of the direction of polarization with two spin 1/2 particles by Gisin and Massar [IV-9], who found that the antiparallel spin pair is more informative than the parallel spin pair. Moreover, Bužek, Derka, and Massar [Chap. 31] applied this method to the estimation of an unknown unitary action.

Further Reading

CHAPTER 22

Optimal Extraction of Information
from Finite Quantum Ensembles

Serge Massar and Sandu Popescu

Abstract. Given only a finite ensemble of identically prepared particles, how precisely can one determine their states? We describe optimal measurement procedures in the case of spin-1/2 particles. Furthermore, we prove that optimal measurement procedures must necessarily view the ensemble as a single composite system rather than as the sum of its components, i.e., optimal measurements cannot be realized by separate measurements on each particle.

A basic assumption of quantum mechanics is that if an infinite ensemble of identically prepared particles is given the quantum state of the particles can be determined exactly. But in practice one never encounters such infinite ensembles, and very often they are not even large (statistical). Given such a finite ensemble its state can only be determined approximately. How much knowledge can be obtained from such finite ensembles? How quickly does one approach exact knowledge as the system becomes large? What experimental strategies furnish the maximum knowledge? We solve these problems in some particular cases. This is not of academic interest only. The solution to such problems is expected to lead to applications in the fields of quantum information transmission, quantum cryptography, and quantum computation.

A fundamental question related to the above has been raised by Peres and Wootters [6]. Is an ensemble of identically prepared particles, viewed as an entity, more than the sum of its components? That is, could more be learned about the ensemble by performing a measurement on all the constituent particles together than by performing separate measurements on each particle? Peres and Wootters conjectured that this is the case. In the present Letter we prove their conjecture, although not in its letter but in its spirit.

We have answered the above problems in the context of a simple “quantum game.” The game consists of many runs. In each run a player receives $N$ spin-$1/2$ particles, all polarized in the same direction. The player knows that the $N$ spins are parallel. He also knows that in each run they are polarized in another direction, randomly and uniformly distributed in space, and that the Hamiltonian of the particles is the same in each run. The player is allowed to do any measurement he wants and is finally required to guess the polarization direction. (The answer must consist of indicating a direction, i.e., the player is not allowed to say something like “with probability $P_1$ the particles were polarized along....”) The score of the run is $\cos^2(\alpha/2)$, where $\alpha$ is the angle between the real and guessed directions. The final score is the average of the scores obtained in each run. The aim is to obtain the maximal score.

As it has been defined the score is a number between 0 and 1. If no measurement is performed, but a polarization direction is simply guessed at randomly, the score is 1/2. The improvement over 1/2 actually represents the “information” gain; scores less than 1/2 also correspond to a gain in information, but in this case the guessed direction is systematically opposite to that of the spins. The maximal score $1$ corresponds to perfect knowledge of the direction of polarization. In this Letter it is shown that the maximum score obtainable is $(N + 1)/(N + 2)$, which tends towards 1 as $N$ tends to infinity, i.e., as expected for infinite ensembles the direction of polarization can be determined exactly.

Clearly, the above “game” is only a particular way in which the problem of optimizing measurements on finite ensembles might be formulated. Peres and Wootters, for example, used the same rules, but a different score, inspired by information theory, and a different distribution of spin directions.

This Letter is organized as follows. A general formalism describing an experiment is used to obtain the equations an optimal experiment must satisfy. Next we consider the game outlined above and obtain $(N + 1)/(N + 2)$ as an upper bound on the score, and optimal experiments that attain this score are exhibited. Last, the conjecture of Peres and Wootters is proven when the system consists of two parallel spins.

Following von Neumann [3], every measurement can be considered as having two stages. The first stage is the interaction between two quantum systems, the measured system and the measuring device; the second stage consists of “reading” the measuring device.

Let us denote by $\{|i\rangle\}$ an orthonormal basis of the Hilbert space of the measured system, and $|\psi_0\rangle_{\text{MD}}$ the initial state of the measuring device.
The first stage of the measurement consists of the interaction between the measured system and the measuring device,
\[ |i⟩ψ_0⟩_{MD} → \sum_f |f⟩|ψ^i_f⟩_{MD}. \]  
(1)

If the measured system started in an arbitrary space \( |ϕ⟩ \) linearity implies
\[ |ϕ⟩ψ_0⟩_{MD} → \sum_{i,f} ⟨i|ϕ⟩|f⟩|ψ^i_f⟩_{MD}. \]  
(2)

To allow for the most general measurements we impose no restrictions on the measuring device nor on the interaction with the measured system. The dimension of the Hilbert space of the measuring device is arbitrary and might be much larger than that of the measured system. The wave functions \( |ψ^i_f⟩_{MD} \) are not necessarily normalized, nor orthogonal to each other. The only constraints they obey are
\[ \sum_f ⟨ψ^i_f|ψ^{i'}_f⟩ = δ_ii', \]  
(3)

which follow from the unitarity of the time evolution describing the interaction with the measured system and the normalization of \( |ψ_0⟩_{MD} \). (To simplify notation, in this formula and throughout the text we drop the subscript MD whenever it is obvious that the state belongs to the Hilbert space of the measuring device.)

The second stage of the experiment consists of reading the state of the measuring device. This is implemented by considering a complete set of orthogonal projectors \( \{P_ξ\} \). Different outcomes of the experiment correspond to finding the measuring device in the different eigenspaces of the projectors \( P_ξ \). The probability of the outcome \( ξ \) if the initial state were \( |ϕ⟩ \) is
\[ P(ϕ, ξ) = \sum_{i,i',f} ⟨ϕ|i'⟩⟨i|ϕ⟩_{MD} ⟨ψ^i_f|P_ξ|ψ^{i'}_f⟩_{MD}. \]  
(4)

Here also, for the sake of generality, the number of possible outcomes \( ξ \) of the measurement is left arbitrary and can be larger than the dimension of the Hilbert space of the measured system.

We stress, because of its importance to our purpose, the complete generality of the about formalism. It includes ideal measurements (as described in the postulates of quantum mechanics [3]) but also fuzzy measurements, repeated experiments on the same system, etc. For example, a positive-operator-valued measure (POVM) [2, 4] is described in our formalism by simply considering the ancilla as a part of the measuring device and letting
the rest of the measuring device act on both the measured system and the ancilla.

Upon finding the measuring device to be in the state $\xi$, some "information" is obtained about the state of the system. This information could be expressed as a function $S(\xi, \phi)$. The average value of $S$ is

$$S = \sum_{\xi} \int \mathcal{D}\phi P(\phi, \xi) S(\phi, \xi),$$

where the sums run over the outcomes $\xi$ of the experiment and the initial states $\phi$ of the system to be measured, with a measure $\mathcal{D}\phi$ corresponding to their distribution.

The problem at hand is to maximize (5) with respect to the possible measurements and guessing strategies, while respecting the unitary relations (3). Below, this program will be carried out in detail in the case of parallel spins.

Before proceeding we simplify the formalism by choosing the $P_\xi$’s to be one dimensional projection operators onto a basis $\{|e_\xi\rangle\}$ of the Hilbert space of the measuring device. Indeed, by decomposing the original projectors as a sum of one dimensional projectors, that is, by a more accurate reading of the measuring device, the information obtained in the measurement can only increase.

We now turn to the specific problem considered in the introduction. The system to be measured consists of $N$ parallel spins polarized in a random direction, say $(\theta, \phi)$. Denote this state

$$|N_{\theta, \phi}\rangle = |\uparrow_{\theta, \phi} \ldots \uparrow_{\theta, \phi}\rangle_N.$$  

(6)

The Hilbert space of the $N$ spins can be decomposed into a sum of subspaces having different total spin $S$ with $S = N/2, N/2 - 1, \ldots$. Since our system consists of $N$ parallel spins, it will always belong to the subspace of highest spin so we have to specify the measuring interaction only for this subspace. A basis of this subspace is $|m\rangle$, $m = -N/2, \ldots, N/2$, which is shorthand for $|S = N/2, S_z = m\rangle$.

The unitary evolution of the spins plus measuring device is given by

$$|m\rangle|\psi_0\rangle_{MD} \rightarrow |v^m\rangle = \sum_{f=1}^{2^N} |f\rangle |\psi_f^m\rangle_{MD},$$  

(7)

where $\{|f\rangle\}$ is a complete base of the Hilbert space of $N$ spins. The prob-
ability to obtain the result $\xi$ is

$$P(N_{\theta, \phi}; \xi) = \sum_{m, m' = -N/2}^{N/2} \sum_{f=1}^{2^N} \langle N_{\theta, \phi} | m \rangle \langle \psi_f^m | e_\xi \rangle \langle e_\xi | \psi_f^{m'} \rangle \langle m' | N_{\theta, \phi} \rangle.$$  \hspace{1cm} (8)

Upon finding the measuring device to be in the state $\xi$, one guesses a direction of polarization $\theta_\xi, \phi_\xi$ and obtains a score $S = S(\theta, \phi; \theta_\xi, \phi_\xi) = \cos^2(\alpha/2)$, as explained in the introduction.

We have finally arrived at the mathematical formulation of our problem. We have to maximize the average score $S_N$,

$$S_N = \sum_{\xi} \int \frac{\sin \theta d\theta d\phi}{4\pi} P(N_{\theta, \phi}; \xi) S(\theta, \phi; \theta_\xi, \phi_\xi),$$  \hspace{1cm} (9)

with the unitary constraints (written in the reading basis $|e_\xi\rangle$)

$$\langle v^m | v^{m'} \rangle = \sum_{\xi, j=1}^{2^N} \sum_{f} \langle \psi_f^m | e_\xi \rangle \langle e_\xi | \psi_f^{m'} \rangle = \delta^{mm'}. $$ \hspace{1cm} (10)

The variables of this problem are $|\psi_f^m\rangle$ which encode the measuring interaction, $|e_\xi\rangle$ which encode the reading procedure, and $\theta_\xi, \phi_\xi$ which encode the guessing strategy. It is worth noting that the final states of the measuring device $|\psi_f^m\rangle$ and the reading base vectors $|e_\xi\rangle$ always appear together, via the scalar product $\langle \psi_f^m | e_\xi \rangle$, so we do not have to vary them independently. Clearly, the reason behind this is that, given a particular measuring interaction and reading procedure, one can always find a completely equivalent experiment by changing both the finite states of the measuring device and the way the result is read.

Rather than work with the complete set of constraints (10), it is convenient to consider first only the constraint

$$\sum_{m = -N/2}^{N/2} \langle v^m | v^m \rangle = \sum_{m = -N/2}^{N/2} \sum_{f} \sum_{\xi} \langle \psi_f^m | e_\xi \rangle \langle e_\xi | \psi_f^m \rangle = N + 1,$$ \hspace{1cm} (11)

which follows immediately from (10). The maximal value of the score obtained by using this single constraint equation (11) is larger or equal to the true maximum, obtained when all the constraints are considered. In our case the two maxima coincide. We shall first find the maximum of the reduced problem and then exhibit a solution of the complete problem that attains the same score.
Upon adding to $S_N$ the constraint (11) multiplied by the Lagrange multiplier $\lambda$ and varying with respect to $\langle \psi^m_f | e_\xi \rangle$ considered as independent variables, one obtains the following linear equations:

$$
\sum_{m'} \langle e_\xi | \psi^{m'}_f \rangle [M_{mm'}(\theta_\xi, \phi_\xi) - \lambda \delta_{mm'}] = 0,
$$

(12)

where

$$
M_{mm'}(\theta_\xi, \phi_\xi) = \int \frac{\sin \theta d\theta d\phi}{4\pi} \langle N_\theta, \phi | m \rangle \langle m | N_\theta, \phi \rangle S(\theta, \phi; \theta_\xi, \phi_\xi).
$$

(13)

Upon multiplying the $m$th equation (12) by $\langle \psi^m_f | e_\xi \rangle$, summing over $m$, $f$, and $\xi$, and using (11), a concise expression for the external value of $S_N$ is found to be $S^\text{extremum}_N = \lambda(N + 1)$.

Equation (12) has a nontrivial solution if and only if $\lambda$ is an eigenvalue of $M(\theta_\xi, \phi_\xi)$ (the trivial solutions correspond to $\langle e_\xi | \psi^m_f \rangle = 0$ for all $m$, implying that the outcome $\xi$ is never realized). The spherical symmetry inherent to this problem can be used to show that the matrix $M(\theta_\xi, \phi_\xi)$ transforms according to the adjoint representation of SU(2):

$$
M(\theta_\xi, \phi_\xi) = U(\theta_\xi, \phi_\xi) M(\theta_\xi = 0) U^\dagger(\theta_\xi, \phi_\xi)
$$

where $U(\theta_\xi, \phi_\xi)$ is an element of the $N + 1$ dimensional irreducible unitary representation of SU(2) that realizes rotations of the spins, sending the $+z$ direction onto the $\theta_\xi, \phi_\xi$ direction. It follows that the eigenvalues of $M(\theta_\xi, \phi_\xi)$ are independent of $|\theta_\xi, \phi_\xi\rangle$. Taking $\theta_\xi = 0$ direct computation shows that $M$ is diagonal and its largest eigenvalue is $\lambda = 1/(N + 2)$. So for this reduced problem the corresponding maximal value of the average score is $S^\text{extremum}_N = (N + 1)/(N + 2)$.

We now exhibit optical experiments that attain the score $S_N = (N + 1)/(N + 2)$, thereby proving that this upper bound on the score can be realized.

In the case $N = 1$, one experiment (among many) that attains the score of $2/3$ is realized by measuring the projection of the spin along a given axis, say the $z$ axis (a Stern-Gerlach experiment), and according to whether the spin is found to be polarized along the $+z$ or $-z$ direction, to guess that this is the direction along which it is polarized.

In the case $N = 2$, one possible optimal experiment consists of the standard measurement of a nondegenerate operator that has the following four eigenstates:

$$
\frac{1}{2} | S \rangle + \frac{\sqrt{3}}{2} \left| \uparrow \hat{n}_i \uparrow \hat{n}_j \right\rangle \quad i = 1, \ldots, 4,
$$

(14)

where $| S \rangle$ is the singlet state, $\uparrow \hat{n}_i$ represents a spin polarized along the $\hat{n}_i$ direction, and the four directions $\hat{n}_i$ are oriented towards the corners of
a tetrahedron. [The phases used in the definition of $\hat{n}_i$ are such that the four states (14) are orthogonal.] The only requirement of the corresponding eigenvalues is that they be different from each other so that the measurement can distinguish between all four eigenstates. If the spins are found to be in the $i$th eigenstate, the guessed direction is $\hat{n}_i$.

In the above optimal measurements the Hilbert space of the measuring device is finite dimensional and the number of possible outcomes of the measurement is finite. By counting the number of parameters it can be shown that such finite dimensional optimal measurements exist in the general case ($N$ spins) but we have not been able to construct one explicitly. However, allowing for an infinite set of possible outcomes and using the spherical symmetry inherent to our problem, a measurement that attains the optimal score $S_{N\text{extremum}}$ can be constructed [5].

The measuring device that gets correlated to the spins is a particle moving on the surface of a sphere. Reading the measuring device consists of measuring the position of the particle. Upon finding it to be located at $\theta, \phi$, one guesses that the spins were aligned along the $\theta, \phi$ direction. Let the reading basis, corresponding to a particle localized at $\theta, \phi$, be denoted $|e_{\theta\phi}\rangle$, with the normalization $\langle e_{\theta\phi}|e_{\theta'\phi'}\rangle = 4\pi\delta_{\theta\theta'}\delta_{\phi\phi'}/\sin \theta$.

The experiment is described by the unitary evolution equation (7) with $|\psi_f^m\rangle$ given by

$$|\psi_f^m\rangle = \sqrt{\frac{N+1}{N}} \sqrt{\frac{1}{2^N}} \int d\theta d\phi \sin \theta \frac{\sin \theta}{4\pi} U_{mN/2}(\theta, \phi)|e_{\theta\phi}\rangle,$$

where $U(\theta, \phi)$ is the same unitary matrix as before. One readily verifies that the $|\psi_f^m\rangle$ obey the unitary relations and that the average score obtained by this measurement is $(N+1)/(N+2)$.

We now come to the crux of our Letter: Must an optimal measurement on an ensemble of parallel spins necessarily treat the ensemble as an entity, i.e., as a single composite system? We show below that there exist no optimal experiments consisting of separate measurements on each spin, even though the result of one measurement may be used to decide which measurement is to be performed next on the other spin. (For related work on optimizing separable measurements, see [1].) For simplicity we consider the case of two spins.

One first makes an arbitrary measurement on the first spin,

$$|\uparrow\rangle_1|\psi_0\rangle_{MD1} \rightarrow |\uparrow\rangle_1|\phi^+\rangle_{MD1} + |\downarrow\rangle_1|\phi^+\rangle_{MD1},$$

$$|\downarrow\rangle_1|\psi_0\rangle_{MD1} \rightarrow |\uparrow\rangle_1|\psi^-\rangle_{MD1} + |\downarrow\rangle_1|\psi^-\rangle_{MD1},$$

(16)
where MD1 denotes the first measuring device. The outcomes of this measurement are obtained by projecting the state of MD1 onto the reading base \(|e_{\xi_1}\rangle\). According to the outcome \(\xi_1\), different measurements are carried out on the second spin (i.e., the interaction of the second measuring device with the second spin is parametrized by \(\xi_1\)),

\[
\begin{align*}
|\uparrow\rangle_2|\psi_0\rangle_{MD2} & \rightarrow |\uparrow\rangle_2|\phi_{\uparrow,\xi_1}^+\rangle_{MD2} + |\downarrow\rangle_2|\phi_{\downarrow,\xi_1}^+\rangle_{MD2}, \\
|\downarrow\rangle_2|\psi_0\rangle_{MD2} & \rightarrow |\uparrow\rangle_2|\psi_{\uparrow,\xi_1}^-\rangle_{MD2} + |\downarrow\rangle_2|\psi_{\downarrow,\xi_1}^-\rangle_{MD2},
\end{align*}
\]

(17)

where MD2 denotes the second measuring device. The outcomes of this measurement are obtained by projecting the state of MD2 onto the reading basis \(|g_{\xi_2,\xi_1}\rangle\); the index \(\xi_1\) appears because the way the results of the second measurement are read may also depend on the outcomes \(\xi_1\) of the first measurement. Putting it all together one obtains

\[
|\uparrow\rangle_1|\uparrow\rangle_2|\varphi_0\rangle_{MD1}|\psi_0\rangle_{MD2} \rightarrow \sum_{f=\uparrow,\downarrow} \sum_{\xi_1,\xi_2} |f\rangle_1|f'\rangle_2 \langle e_{\xi_1}|\psi_f^+\rangle \langle g_{\xi_2,\xi_1}|e_{\xi_1}\rangle |g_{\xi_2,\xi_1}\rangle
\]

(18)

and similarly for the other initial states \(|\uparrow\rangle_1|\downarrow\rangle_2\), \(|\downarrow\rangle_1|\uparrow\rangle_2\), \(|\downarrow\rangle_1|\downarrow\rangle_2\). Equation (18) is a particular case of the general evolution (7), the measuring basis \(|e_{\xi}\rangle\) being replaced by the basis \(|e_{\xi_1}\rangle|g_{\xi_2,\xi_1}\rangle\). Indeed the two successive measurements considered here correspond in the general formalism to a single measuring device consisting of the two pieces MD1 and MD2, and the action of the human observer who “reads” the result of MD1 and decides accordingly what measurement to do next is replaced by MD2 automatically getting correlated to the final state of MD1, and tuning its interaction with the second spin accordingly.

The unitary relations (10) are now replaced by the unitary relations obeyed by each measuring device separately:

\[
\begin{align*}
\sum_{f,\xi_1} a_{f,\xi_1}^+ (a_{f,\xi_1}^-)^* & = 1, & \sum_{f,\xi_1} a_{f,\xi_1}^- (a_{f,\xi_1}^+)^* & = 1, \\
\sum_{f,\xi_1} a_{f,\xi_1}^+ (a_{f,\xi_1}^-)^* & = 0, & \sum_{f,\xi_1} a_{f,\xi_1}^- (a_{f,\xi_1}^+)^* & = 0,
\end{align*}
\]

(19)

\[
\begin{align*}
\sum_{f',\xi_2} b_{f',\xi_2,\xi_1}^+ (b_{f',\xi_2,\xi_1}^-)^* & = 1, & \sum_{f',\xi_2} b_{f',\xi_2,\xi_1}^- (b_{f',\xi_2,\xi_1}^+)^* & = 1, \\
\sum_{f',\xi_2} b_{f',\xi_2,\xi_1}^+ (b_{f',\xi_2,\xi_1}^-)^* & = 0, & \sum_{f',\xi_2} b_{f',\xi_2,\xi_1}^- (b_{f',\xi_2,\xi_1}^+)^* & = 0,
\end{align*}
\]

(20)

where, as above, \(f, f' = \uparrow, \downarrow\) and

\[
\begin{align*}
a_{f,\xi_1}^+ & = \langle e_{\xi_1}|\psi_f^+\rangle, & b_{f',\xi_2,\xi_1}^+ & = \langle g_{\xi_2,\xi_1}|\varphi_{f',\xi_1}^+\rangle, \\
a_{f,\xi_1}^- & = \langle e_{\xi_1}|\psi_f^-\rangle, & b_{f',\xi_2,\xi_1}^- & = \langle g_{\xi_2,\xi_1}|a\varphi_{f',\xi_1}^-\rangle.
\end{align*}
\]

(21)
After completing these two measurements, one guesses a direction of polarization $\theta_{\xi_1,\xi_2}, \phi_{\xi_1,\xi_2}$ which depends of course on both outcomes.

If (18) is to describe an optimal experiment it must satisfy (12) with $\lambda = 1/4$ [since any optimal experiment on two parallel spins must necessarily also be an extremum of the reduced problem with (11) as the only constraint]. Explicitly Eq. (12) takes the form (dropping the indices $f, f', \xi_1, \xi_2$)

$$-2Sa^+b^+ + CE(a^+b^- + a^-b^+) = 0,$$

$$2CSE^*a^+b^+ - (a^+b^- + a^-b^+) + 2CSEa^-b^- = 0,$$

$$SE^*(a^+b^- + a^-b^+) - 2Ca^-b^- = 0,$$

(22)

which must be satisfied for all $i, j, \xi_1, \xi_2$. We have used the notation $C = \cos(\theta_{\xi_1,\xi_2}/2), S = \sin(\theta_{\xi_1,\xi_2}/2)$ and $E = e^{i\phi_{\xi_1,\xi_2}}$. Equations (22) solve to yield

$$\frac{a^+_{f,\xi_1}}{a^-_{f,\xi_1}} = \frac{b^+_{f',\xi_2,\xi_1}}{b^-_{f',\xi_2,\xi_1}} = \frac{CE}{S}. \quad (23)$$

Upon inserting this relation into the unitary relations (20) a contradiction is readily obtained, thereby proving that experiments such as (18) cannot be optimal experiments.

We have generalized this proof to the case where a finite number of measurements are carried out alternatively on the two spins. Whether an infinite number of such alternating measurements can reach the optimal score is still an open problem.

It is a pleasure for us to thank P. Bieliavsky for very helpful discussions.

References

CHAPTER 23

Asymptotic Estimation Theory
for a Finite-Dimensional Pure State Model

Masahito Hayashi

Abstract. The optimization of measurement for \( n \) samples of pure states are studied. The error of the optimal measurement for \( n \) samples is asymptotically compared with the one of the maximum likelihood estimators from \( n \) data given by the optimal measurement for one sample.

1. Introduction

Recently, there has been a rise in the necessity for studies about statistical estimation for the unknown state, related to the corresponding advance in measuring technologies in quantum optics. An investigation including both quantum theory and mathematical statistics is necessary for an essential understanding of quantum theory because it has statistical aspects [15, 17]. Therefore, it is indeed important to optimize the measuring process with respect to the estimation of the unknown state. Such research is known as quantum estimation, and was initiated by Helstrom in the late 1960s, originating in the optimization of the detecting process in optical communications [15]. In the classical statistical estimation, one searches for the most suitable estimator for which probability measure describes the objective probabilistic phenomenon. In quantum estimation, one searches for the most suitable measurement for which density operator describes the objective quantum state.

Contained among important results are three estimation problems. The first is of the complex amplitude of coherent light in thermal noise and the second is of the expectation parameters of quantum Gaussian state. The former was studied by Yuen and Lax [25] and the latter by Holevo [17]. These studies discovered that heterodyning is the most suitable for the estimation of the complex amplitude of coherent light in thermal noise.

This chapter is reprinted from *J. Phys. A: Math. and Gen.*, 31, 4633–4655, 1998. Since Appendices consist of the technical proofs, they are omitted. They are available at quant-ph/9704041.
The third is a formulation of the covariant measurement with respect to an action of a group. It was studied by Holevo [16, 17]. In the formulation, he established a quantum analogue of Hunt-Stein theorem.

Quantum estimation, was first used in the evaluation of the estimation error of a single sample of the unknown state as it had advanced in connection with the optimization of the measuring process in optical communications. Thus, early studies were lacking in asymptotic aspects, i.e., there were little researches with respect to reducing the estimation error by quantum correlations between samples.

Recently, studies concerning the estimation of the unknown state have attracted many physicists [4, 5, 18, 19]. Some of them were drawn by the variation of the measuring precision with respect to the number of samples of the unknown state [2, 24].

Nagaoka [23] studied, for the first time, asymptotic aspects of quantum estimation. He paid particular attention to the quantum correlations between samples of the unknown state, and studied the relation between the asymptotic estimation and the local detection of a one-parameter family of quantum states.

In the early 1990s, Fujiwara and Nagaoka [8, 9, 10] studied the estimation problem for a multiparameter family consisting of pure states. They pioneered studies into the estimation problem of the complex amplitude of noiseless coherent light. However, there had been some studies with respect to that of coherent light in thermal noise. The research found that heterodyning is the most suitable for the estimation of the complex amplitude of noiseless coherent light. In 1996, Matsumoto [21] established a more general formulation of the estimation for a multiparameter family consisting of pure states. Moreover in 1991, Nagaoka [22] treated the estimation problem for two-parameter families of mixed states in a spin-$\frac{1}{2}$ system, and in 1997 the author [12, 13] treated it for 3-parameter families of mixed states in a spin-$\frac{1}{2}$ system. However, there are no asymptotic aspects in these works concerning multiparameter families. There is more need of this type of investigation into one- and multiparameter families.

Can quantum estimation reduce the estimation error by using the quantum correlations between samples, under the preparation of sufficient samples of the unknown state? To answer this question, in this paper, we treat a family, consisting of all of pure states on a Hilbert space $\mathcal{H}$ under the

*Where $\mathcal{H}$ denotes a finite-dimensional Hilbert space which corresponds to the physical system of interest.*
preparation of \( n \) samples of the unknown state, with the estimation problem. In section 2, we use, as a tool, the composite system consisting of \( n \) samples as a single system. The quantum i.i.d. condition is introduced as the quantum counterpart of the independent and identical distributions condition (3). In section 3, we review Holevo’s result concerning covariant measurements which will be used in the following sections. In section 4, we apply Holevo’s result to the optimization of measurements on the composite system, which results in obtaining the most suitable measurement (theorem 6). We asymptotically calculate the estimation error by the optimal measurement in the sense of both the error mean square and large deviation (see (9), (10), (11), and (13)). The first term of the right-hand side of (10) is consistent with the value conjectured from the results by Fujiwara and Nagaoka [10] and Matsumoto [21]. However, the optimal measurement may be too difficult for modern technology to realize when using more than one sample.

In section 5, we use this estimation problem under the following guidelines. The samples are divided into pairs consisting of a maximum of \( m \) samples. By measuring each pair with the optimal measurement in section 4, we create some data. The estimated value is given by manipulating these data. The restricted condition is called \( m \)-semiclassical (see (14)). We compare an \( m \)-semiclassical measurement with the optimal measurement of section 4, with respect to the estimation error under the preparation of a sufficient amount of samples. When we use the maximum likelihood estimator to manipulate the data, the error mean square of both asymptotically coincide in the first order (see (10) and (19)). However, when the radius of allowable errors is finite, the error of large deviation type in the latter is smaller than that in the former type (see (11) and (20)). Both coincide in the case of the maximum likelihood estimator under the limit where the radius goes to infinitesimal (13), (18). Can we asymptotically realize a small estimation error as the optimal measurements in section 4 has? It is, physically, sufficient to construct the optimal measurement for one sample. In section 5, we show how to construct it (see (25)).

Most of the proofs of this paper are given in the appendices. In view of multiparameter families of mixed states in spin-\( \frac{1}{2} \) system, Hayashi [14] has discussed the same problem using Cramér-Rao type bound.
2. Pure State $n$-I.I.D. Model

In this section, we use the mathematical formulation of the estimation for pure states. Let $k$ be the dimension of the Hilbert space $\mathcal{H}$, and $\mathcal{P}(\mathcal{H})$ be the set of pure states on $\mathcal{H}$.

In quantum physics, the most general description of a quantum measurement is given by the mathematical concept of a positive operator-valued measure (POVM) [15, 17] on the system of state space. Generally, if $\Omega$ is measurable space, a measurement $M$ satisfies the following

$$M(B) = M(B)^* \quad M(B) \geq 0 \quad M(\phi) = 0 \quad M(\Omega) = \text{Id on } \mathcal{H} \quad \forall B \subset \Omega$$

$$M(\cup_i B_i) = \sum_i M(B_i) \quad \text{for } B_i \cap B_j = \phi (i \neq j)$$

$\{B_i\}$ is countable subsets of $\Omega$.

In this paper, $\mathcal{M}(\Omega, \mathcal{H})$ denotes the set of POVMs on $\mathcal{H}$ whose measurable set is $\Omega$. A measurement $M \in \mathcal{M}(\Omega, \mathcal{H})$ is said to be simple if $M(B)$ is a projection for any Borel $B \subset \Omega$. A measurement $M$ is random if it is described as a convex combination of simple measurements. A random measurement $M = \sum_i a_i M_i$ ($M_i$ is simple and $a_i > 0$.) can be realized when every measurement $M_i$ is taken with the probability $a_i$.

In this paper, we consider measurements whose measurable set is $\mathcal{P}(\mathcal{H})$ since it is known that the unknown state is included in $\mathcal{P}(\mathcal{H})$.

Next, we define two distances charactering the homogeneous space $\mathcal{P}(\mathcal{H})$.

**Definition 1:** Fubini-Study distance $d_{fs}$ (which is the geodesic distance of the Fubini-Study metric) is defined as:

$$\cos d_{fs}(\rho, \hat{\rho}) = \sqrt{\text{tr} \rho \hat{\rho}} \quad 0 \leq d_{fs}(\rho, \hat{\rho}) \leq \frac{\pi}{2}. \quad (1)$$

The Bures’ distance $d_b$ is defined in the usual way:

$$d_b(\rho, \hat{\rho}) := \sqrt{1 - \text{tr} \rho \hat{\rho}}. \quad (2)$$

It is introduced by Bures [3] in a mathematical context.

Let $W(\rho, \hat{\rho})$ be a measure of deviation of the measured value $\hat{\rho}$ from the actual value $\rho$, then we have the following equivalent conditions.

- $W(\rho, \hat{\rho}) = W(g\rho g^*, g\hat{\rho} g^*) \forall \rho, \forall \hat{\rho} \in \mathcal{P}(\mathcal{H}) \forall g \in \text{SU}(k)$.
- There exists a function $h$ on $[0, 1]$ such that $W(\rho, \hat{\rho}) = h \circ d_{fs}(\rho, \hat{\rho})$.

It is natural to assume that a deviation measure $W(\rho, \hat{\rho})$ is monotone increasing with respect to the Fubini-Study distance $d_{fs}$.
If $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are $n$ Hilbert spaces which correspond to the physical systems, then their composite system is represented by the tensor Hilbert space $\mathcal{H}^{(n)} := \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n = \otimes_{i=1}^n \mathcal{H}_i$. Thus, a state on the composite system is denoted by a density operator $\rho$ on $\mathcal{H}^{(n)}$. In particular if $n$ element systems $\{\mathcal{H}_i\}$ of the composite system $\mathcal{H}^{(n)}$ are independent of each other, there exists a density $\rho_i$ on $\mathcal{H}_i$ such that $\rho^{(n)} = \rho_1 \otimes \cdots \otimes \rho_n = \otimes_{i=1}^n \rho_i$. The condition:

$$\mathcal{H}_1 = \cdots = \mathcal{H}_n = \mathcal{H}, \quad \rho_1 = \cdots = \rho_n = \rho$$

(3)
corresponds to the independent and identically distributed (i.i.d.) condition in the classical case. In this paper, we use this estimation problem under this condition (3) called the quantum i.i.d. condition. This condition means that identical $n$ samples are independently prepared. The model $\{\rho^{(n)} = \rho \otimes \cdots \otimes \rho | \rho \in \mathcal{P}(\mathcal{H})\}$ is called $n$-i.i.d. model. As $\rho$ is a pure state, $\mathcal{H}^{(n)}$ and $\rho^{(n)}$ are simplified as follows. Letting $\rho = |\phi\rangle \langle \phi| \in \mathcal{P}(\mathcal{H})$, we have $\rho^{(n)} = |\phi^{(n)}\rangle \langle \phi^{(n)}|$, $\phi^{(n)} := \phi \otimes \cdots \otimes \phi$. Because all of the vectors $\phi^{(n)}$ is included in $n$-times symmetric tensor space, for any measurement $M \in \mathcal{M}(\Omega, \mathcal{H}^{(n)})$ on the $n$-times tensor space $\mathcal{H}^{(n)}$, the measurement $\tilde{M}(d\omega) := P_{\mathcal{H}_s^{(n)}} M(d\omega) P_{\mathcal{H}_s^{(n)}} \in \mathcal{M}(\Omega, \mathcal{H}_s^{(n)})$ on the $n$-times symmetric tensor space $\mathcal{H}_s^{(n)}$ satisfies that $\text{tr} M(d\omega) \rho^{(n)} = \text{tr} \tilde{M}(d\omega) \rho^{(n)}$ for any $\rho \in \mathcal{H}$, where $\mathcal{H}_s^{(n)}$ denotes the $n$-times symmetric tensor space on $\mathcal{H}$. Therefore, all possible measurements can be regarded as elements of $\mathcal{M}(\mathcal{P}(\mathcal{H}), \mathcal{H}_s^{(n)})$. The mean error of the measurement $\Pi \in \mathcal{M}(\mathcal{P}(\mathcal{H}), \mathcal{H}_s^{(n)})$ with respect to a deviation measure $W(\rho, \hat{\rho})$, provided that the actual state is $\rho$, is equal to $D^W_{\rho}(\Pi) := \int_{\mathcal{P}(\mathcal{H})} W(\rho, \hat{\rho}) \text{tr}(\Pi(\hat{\rho}) \rho^{(n)})$. In minimax approach the maximum possible error with respect to a deviation measure $W(\rho, \hat{\rho})$, $D^W_{\rho}(\Pi) := \max_{\rho \in \mathcal{P}(\mathcal{H})} D^W_{\rho}(\Pi)$ is minimized.

3. Quantum Hunt-Stein Theorem

In this section, the quantum Hunt-Stein theorem, established by Holevo [17, 16], is summarized. Let $G$ be a compact transitive Lie group of all transformations on a compact parametric set $\Theta$, and $\{V_g\}$ a continuous unitary irreducible representation of $G$ in a finite-dimensional Hilbert space $\mathcal{H}^\prime := \mathbb{C}^k$, and $\mu$ a $\sigma$-finite invariant measure on group $G$ such that $\mu(G) = 1$. In this section, we consider the following condition for a measurement.

**Definition 2:** A measurement $\Pi \in \mathcal{M}(\Theta, \mathcal{H}^\prime)$ is covariant with respect to
\{V_g\}$ if $V_g^*\Pi(B)V_g = \Pi(B_{g^{-1}})$ for any $g \in G$ and any Borel $B \subset \Theta$, where $B_g := \{g\theta|\theta \in B\}$. $\mathcal{M}(\Theta, V)$ denotes the set of covariant measurements with respect to $\{V_g\}$.

Covariant measurements are characterized by the following theorem.

**Theorem 3:** The map $V^\theta$ from the set $S(\mathcal{H}')$ of densities on $\mathcal{H}'$ to $\mathcal{M}(\Theta, V)$ is surjective for any $\theta \in \Theta$, where $V^\theta(P) := k' \int_{\{g\theta\in B\}} V_g PV^*_g \mu(dg)$ for any $B \in \mathcal{B}(\Theta)$ and $P \in S(\mathcal{H}')$.

In this section, we treat with the following condition for a family of states.

**Definition 4:** The family is called covariant under the representation $\{V_g\}$ of group $G$ acting on $\Theta$, if $S_{g\theta} = V_g S_\theta V^*_g$ for any $g \in G$ and $\theta \in \Theta$.

Assuming that the object is prepared in one of the states $\{S_\theta|\theta \in \Theta\}$ but the actual value of $\theta$ is unknown, then the difficulty is estimating this value as close as possible to a measurement on the object. We shall solve this problem by means of the quantum statistical decision theory.

Let $W(\theta, \hat{\theta})$ be a measure of deviation of the measured value $\hat{\theta}$ from the actual value $\theta$. It is natural to assume that $W(\theta, \hat{\theta})$ is invariant:

$$W(\theta, \hat{\theta}) = W(g\theta, g\hat{\theta}) \quad \forall \theta, \forall \hat{\theta} \in \Theta, \forall g \in G. \quad (4)$$

The mean error of the measurement $\Pi \in \mathcal{M}(\Theta, \mathcal{H}')$ with respect to a deviation measure $W(\theta, \hat{\theta})$, provided that the actual state is $S_\theta$, is equal to $D_{W,S}^\theta(\Pi) := \int_{\Theta} W(\theta, \hat{\theta}) \text{tr}(\Pi(d\hat{\theta})S_\theta)$. Following the classical statistical decision theory, we can form two functionals of $D_{W,S}^\theta$ giving a total measure of precision of the measurement $\Pi$.

In Bayes’ approach we take the mean of $D_{W,S}^\theta$ with respect to a given prior distribution $\pi(d\theta)$. The measurement minimizing the resulting functional $D_{W,S}^\pi(\Pi) := \int_{\Theta} D_{W,S}^\pi(\Pi) \pi(d\theta)$ is called Bayesian. This quantity represents the mean error in the situation where $\theta$ is a random parameter with known distribution $\pi(d\theta)$. In particular, as $\Theta, G$ are compact and ‘nothing is known’ about $\theta$, it is natural to take for $\pi(d\theta)$ the ‘uniform’ distribution, i.e., normalized invariant measure $\nu(d\theta)$ defined as $\nu(B) := \mu(\{g\theta \in B\})$. It is independent of the choice of $\theta \in \Theta$.

In minimax approach the maximum possible error with respect to a deviation measure $W(\theta, \hat{\theta})$, $D_{W,S}^W(\Pi) := \max_{\theta \in \Theta} D_{W,S}^\theta(\Pi)$ is minimized. The minimizing measurement is called minimax.

Because $G$ is compact, we shall show that in the covariant case the minima of Bayes and minimax criteria coincide and are achieved on a covari-
ant measurement. We obtain the following quantum Hunt-Stein theorem [16, 17]. It is easy to prove the theorem.

**Theorem 5:** For a covariant measurement $\Pi \in M(\Theta, V)$, we obtain $D^W_S(\Pi) = D^W_S(\Pi) = D^W(\Pi)$.

For $\Pi \in M(\Theta, \mathcal{H}')$, denote $\Pi_g(B) := V_g\Pi(B_g)V_g^*$ for $B \in \mathcal{B}(\Theta)$. Introducing the “averaged” measurement $\Pi(B) := \int_G \Pi_{g^{-1}}(B)\mu(dg)$, we have $D^W_S(\Pi) = \int_G D^W_S(\Pi_{g^{-1}})\mu(dg) = D^W_S(\Pi)$. Thus, $D^W_S(\Pi) \geq D^W_S(\Pi) = D^W_S(\Pi)$. In this case, minimax approach and Bayes’ approach with respect to $\nu(d\theta)$ are equivalent. Therefore we minimize the following value $D^W_S(\nu) \circ V^\theta(P) = k' \int_G W(\theta, g\theta) tr\ S_\theta V_g PV_g^* \mu(dg) = tr\hat{W}(\theta)P$, where

$$\hat{W}(\theta) := k' \int_G W(\theta, g\theta)V_g^* S_\theta V_g \mu(dg) = k' \int_\Theta W(\theta, \hat{\theta})S_\hat{\theta} \nu(d\hat{\theta})$$

Thus, it is sufficient to consider the following minimization $\min_{P \in S(\mathcal{H})} tr\hat{W}(\theta)P = \min_{P \in P(\mathcal{H}')} tr\hat{W}(\theta)P$.

### 4. Optimal Measurement in Pure State $n$-I.I.D. Model

In this section we apply the theory of section 3 to the problem of section 2.

We let $\Theta := P(\mathcal{H})$, $\mathcal{H}' := \mathcal{H}^{(n)}$, $G := SU(k)$, and $S_\rho := \rho^{(n)}$. Then, the invariant measure $\nu$ on $P(\mathcal{H})$ is equivalent to the measure defined by the volume bundle induced by the Fubini-Study metric. We let the action $\{V_g\}$ of $G = SU(k)$ to $\mathcal{H}^{(n)}$ be the tensor representation of the natural representation. In this case, we have $k' = \binom{n+k-1}{k-1}$.

**Theorem 6:** If a deviation measure $W(\rho, \hat{\rho})$ is monotone increasing with respect to the Fubini-Study distance $d_I$, we find that $\min_{P_0 \in P(\mathcal{H}^{(n)})} tr\hat{W}(\rho)P_0 = tr\hat{W}(\rho)\rho^{(n)}$.

For a proof see appendix A. Thus, $V^\rho(\rho^{(n)})$ is the optimal measurement with respect to a deviation measure $W(\rho, \hat{\rho})$. The optimal measurement is independent of the choice of $\rho$ and $W$ since $V^\rho_0(\rho^{(n)}_0) = V^\rho(\rho^{(n)}).$ This optimal measurement is denoted by $\Pi_n$ and is described as $\Pi_n(d\hat{\rho}) := \binom{n+k-1}{k-1} \hat{\rho}^{(n)} \nu(d\hat{\rho})$. Under the following chart (6), the optimal measurements are denoted as:

$$\Pi_n(d\theta) = \binom{n+k-1}{k-1} |\phi(\theta)^{(n)}\rangle \langle \phi(\theta)^{(n)}| \nu(d\theta) \quad (5)$$
for $\theta \in \{\theta \in \mathbb{R}^{2k-2}|\theta_1, \ldots, \theta_{k-1} \in [0, \pi/2], \theta_k, \ldots, \theta_{2k-2} \in [0, 2\pi]\}$, where we defined as follows:

$$
\phi(\theta) := \begin{pmatrix}
\cos \theta_1 \\
e^{i\theta_k} \sin \theta_1 \cos \theta_2 \\
e^{i\theta_{k+1}} \sin \theta_1 \sin \theta_2 \cos \theta_3 \\
\vdots \\
e^{i\theta_{2k-3}} \sin \theta_1 \sin \theta_2 \sin \theta_3 \cdots \sin \theta_{k-2} \cos \theta_{k-1} \\
e^{i\theta_{2k-2}} \sin \theta_1 \sin \theta_2 \sin \theta_3 \cdots \sin \theta_{k-2} \sin \theta_{k-1}
\end{pmatrix}.
$$

(6)

The invariant measures $\nu(\, d\theta)$ described above is from p. 31 [11]

$$
\nu(\, d\theta) = \frac{(k - 1)!}{\pi^{k-1}} \sin^{2k-3} \theta_1 \sin^{2k-5} \theta_2 \cdots \sin \theta_{k-1} \cos \theta_1 \cos \theta_2 \cdots \cos \theta_{k-1} \, d\theta_1 \, d\theta_2 \cdots d\theta_{2k-2}.
$$

(7)

**Lemma 7:** If the deviation measure $W$ is characterized as $W(\rho, \hat{\rho}) = h \circ d_{fs}(\rho, \hat{\rho})$, we can describe the maximum possible error of the optimal measurement $\Pi_n$ as $D_{W, (n)}(\Pi_n) = 2(k - 1)\binom{n + k - 1}{k-1} \int_0^{\frac{\pi}{2}} h(\theta) \cos^{2n+1} \theta \sin^{2k-3} \theta \, d\theta$.

For a proof, see appendix B.

Next, we asymptotically calculate the error of the optimal measurements $\Pi_n$ in the third order.

**Theorem 8:** When the deviation measure $W$ is described as $W = d^\gamma_b$, we can asymptotically calculate the maximum possible error of the optimal measurement as:

$$
\lim_{n \to \infty} D_{d^\gamma_b, (n)}(\Pi_n) n^{\frac{\gamma}{2}} = \frac{\Gamma(k - 1 + \gamma/2)}{\Gamma(k - 1)}.
$$

(8)

Specially in the case of $\gamma = 2$, we have

$$
D_{d^2_b, (n)}(\Pi_n) n = \frac{(k - 1)n}{n + k} = (k - 1) \sum_{i=0}^{\infty} \left(-\frac{k}{n}\right)^i \to k - 1 \text{ as } n \to \infty.
$$

(9)

When the deviation measure is defined by the square of the Fubini-Study distance, we can asymptotically calculate the maximum possible error of the optimal measurement as:

$$
D_{d^2_{fs}, (n)}(\Pi_n) n \cong (k - 1) - \frac{2}{3} k(k - 1) \frac{1}{n} + k(k - 1) \frac{23k - 7}{45} \frac{1}{n^2}
$$

as $n \to \infty$.

(10)
The error of the sequence \( \{\Pi_n\}_{n=1}^{\infty} \) of the optimal measurements can be calculated in the sense of large deviation as:

\[
\frac{1}{n} \log \left( \Pr_{\Pi_n}^{\rho(n)} \{ \hat{\rho} \in \mathcal{P}(\mathcal{H}) | d_{fs}(\rho, \hat{\rho}) \geq \varepsilon \} \right)
\]

\[
\cong \log \cos^2 \varepsilon + (k - 2) \frac{\log n}{n}
\]

\[
+ \left( - \log(k - 2)! + 2(k - 2) \log(\sin \varepsilon) - 2 \log(\cos \varepsilon) \right) \frac{1}{n}
\]

\[
+ \left( \frac{k^2 - k - 2}{2} + (k - 2) \cot^2 \varepsilon \right) \frac{1}{n^2}
\]

as \( n \to \infty \), \hspace{1cm} (11)

where \( \Pr_M^S B \) denotes the probability of \( B \) with respect to the probability measure \( \text{tr}(M(d\omega)S) \) for a Borel \( B \subset \Omega \), a measurement \( M \in \mathcal{M}(\Omega, \mathcal{H}') \) and a state \( S \in \mathcal{S}(\mathcal{H}') \).

For a proof, see appendix C. The first term of the right hand of (11) coincide with the logarithm of the fidelity [20].

In this paper, \( \varepsilon \) in equations (11) is called admissible radius.

Since

\[
\lim_{\varepsilon \to 0} \frac{\log \cos^2 \varepsilon}{\varepsilon^2} = -1
\]

we obtain the following large deviation approximation

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log \left( \Pr_{\Pi_n}^{\rho(n)} \{ \hat{\rho} \in \mathcal{P}(\mathcal{H}) | d_{fs}(\rho, \hat{\rho}) \geq \varepsilon \} \right) = -1.
\]

5. Semiclassical Measurement

In this section, we consider measurements which allow the quantum correlation between finite samples only. A measurement \( M \) on \( \mathcal{H}^{(nm)} \) is called \( m \)-semiclassical if there exists an estimator \( T \) on the probability space \( \mathcal{P}(\mathcal{H}) \times \cdots \times \mathcal{P}(\mathcal{H}) \) whose range is \( \mathcal{P}(\mathcal{H}) \) such that

\[
M(B) = \int_{T^{-1}(B)} \Pi_m(d\rho_1) \otimes \cdots \otimes \Pi_m(d\rho_n) \quad \forall B \subset \mathcal{P}(\mathcal{H}).
\]

We compare the error between \( m \)-semiclassical measurements and the optimal measurement \( \Pi_{nm} \) for \( nm \) samples of the unknown state as the equations (10), (11), and (13).

In doing this comparison, we bear in mind asymptotic estimation theory in classical statistics. In classical statistics, it is assumed that the sequence of estimators satisfies the consistency.
Definition 9: A sequence \( \{T^{(n)}\}_{n=1}^{\infty} \) of estimators on a probability space \( \Omega \) is called consistent with respect to a family \( \{p_{\theta} | \theta \in \Theta \} \) of probability distributions on \( \Omega \), if it satisfies the condition (15), where every \( T^{(n)} \) is a probability variable on the probability space \( \bigotimes_{n} \Omega \) whose domain is \( \Theta \)

\[
p_{\theta}^{(n)} \{d_{J}(T^{(n)}, \theta) > \varepsilon \} \to 0 \quad \text{as} \quad n \to \infty \quad \forall \theta \in \Theta, \forall \varepsilon > 0 \tag{15}
\]

where \( d_{J} \) denotes the geodesic distance defined by the Fisher information metric and \( p_{\theta}^{(n)} \) denotes the probability measure \( p_{\theta} \times \cdots \times p_{\theta} \) on the probability space \( \bigotimes_{n} \Omega \).

It is well known that the following theorem establishes under the preceding consistency [1, 6, 7].

Theorem 10: If a sequence \( \{T^{(n)}\}_{n=1}^{\infty} \) of estimators is a consistent estimator with respect to a family \( \{p_{\theta} | \theta \in \Theta \} \) of probability distributions on a probability space \( \Omega \) which satisfies some regularity, then we have the following inequalities

\[
\lim_{n \to \infty} n \int \ldots \int d_{J}^{2}(T^{(n)}(x_{1}, x_{2}, \ldots, x_{n}), \theta) p_{\theta}^{(n)}(dx_{1}, \ldots, dx_{n}) \geq \dim \Theta \tag{16}
\]

\[
\lim_{n \to \infty} \frac{1}{n} \log \left( p_{\theta}^{(n)} \{D(p_{T^{(n)}} || p_{\theta}) \geq \varepsilon \} \right) \geq -\varepsilon \tag{17}
\]

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\varepsilon^{2} n} \log \left( p_{\theta}^{(n)} \{d_{J}(T^{(n)}, \theta) \geq \varepsilon \} \right) \geq -\frac{1}{2} \tag{18}
\]

where \( D(p||q) \) denotes the information divergence of a probability distribution \( q \) with respect to another probability distribution \( p \) defined by

\[
D(p||q) \ := \ \int_{\Omega} (\log p(\omega) - \log q(\omega)) p(\omega) d\omega
\]

Under some regularity conditions, the lower bounds of (16) and (18) can be attained by the maximum likelihood estimator when the family \( \{p_{\theta} | \theta \in \Theta \} \) is exponential, but generally cannot be attained.

For the comparison, we apply theorem 10 to the family of distributions \( \{\text{tr} \Pi_{m}(d \hat{\rho}) \rho^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\} \) given by the measurement \( \Pi_{m} \) and the family of states \( \{\rho^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\} \). We consider the sequence of measurements \( \{T_{(n,m)}\}_{n=1}^{\infty} \) which corresponds to the consistent estimator \( \{T^{(n)}\}_{n=1}^{\infty} \) on the family of distributions \( \{\text{tr} \Pi_{m}(d \hat{\rho}) \rho_{1}^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\} \), where \( T_{(n,m)} \) is the
measurement on $\mathcal{H}^{(nm)}$ defined by the estimator $T^{(n)}$ and $n$ data given by the measurement $\Pi_m \otimes \cdots \otimes \Pi_m$ and the state $\rho^{(nm)}$. From the symmetry of $\mathcal{P}(\mathcal{H})$ and $\Pi_m$, the information divergence of a probability measure with respect to another a probability measure $\text{tr} \Pi_m (d\hat{\rho}) \rho^{(m)}_2$ is determined by the the Fubini-Study distance $\varepsilon$ between $\rho_1$ and $\rho_2$. Thus, the divergence can be denoted by $D_m(\varepsilon)$. From lemma 11, the geodesic distance $d_{\Pi_m}$ with respect to Fisher information metric in the family of distributions $\{\text{tr} \Pi_m (d\hat{\rho}) \rho^{(m)} | \rho \in \mathcal{P}(\mathcal{H})\}$ is given by $d_{\Pi_m} = \sqrt{2md_{fs}}$. Since $\dim \mathcal{P}(\mathcal{H}) = 2(k-1)$, we have the following inequalities:

$$
\lim_{n \to \infty} \frac{nm}{\lim_{n \to \infty}} D^2_{fs,(nm)}(T_{(n,m)}) = \\
\lim_{n \to \infty} \max \rho \in \mathcal{P}(\mathcal{H}) nm \int \rho \text{tr}(T_{(n,m)}(d\hat{\rho})\rho^{(nm)}) \geq k - 1 \quad (19)
$$

$$
\lim_{n \to \infty} \frac{1}{nm} \log \text{Pr}^{(nm)}_{T_{(n,m)}} \{ \hat{\rho} \in \mathcal{P}(\mathcal{H}) | d_{fs}\rho, \hat{\rho} \geq \varepsilon \} \geq -\frac{D_m(\varepsilon)}{m} \quad (20)
$$

$$
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 nm} \log \text{Pr}^{(nm)}_{T_{(n,m)}} \{ \hat{\rho} \in \mathcal{P}(\mathcal{H}) | d_{fs}\rho, \hat{\rho} \geq \varepsilon \} \geq -1. \quad (21)
$$

The lower bound of (19) is consistent with the first term of the right-hand side of (10) and the lower bound of (21) is consistent with the right-hand side of (13). When the sequence of measurements $\{T^{(n)}\}_{n=1}^\infty$ corresponds to the maximum likelihood estimator, the lower bounds (19) and (21) can be attained. We have the following lemma concerning the comparison of the lower bound $-\frac{D_m(\varepsilon)}{m}$ of (20) and the first term $2 \log \cos \varepsilon$ of the right-hand side of (11).

**Lemma 11:** We can calculate the divergence $D_m(\varepsilon)$ and the distance $d_{\Pi_m}$ as:

$$
\frac{D_m(\varepsilon)}{m} = \sum_{i=1}^{m} \frac{\sin^2 i \varepsilon}{i} \to - \log (1 - \sin^2 \varepsilon) = - \log \cos^2 \varepsilon \quad \text{as} \ m \to \infty \quad (22)
$$

$$
d_{\Pi_m} = \sqrt{2md_{fs}}. \quad (23)
$$

Therefore, $\frac{D_m(\varepsilon)}{m}$ is monotone increasing with respect to $m$.

For a proof, see appendix D. (22) infers that

$$
0 < \frac{-m \log \cos^2 \varepsilon - D_m(\varepsilon)}{m \varepsilon^{2m}} \to 0 \quad \text{as} \ \varepsilon \to 0 \quad (24)
$$

which means that the first term on the right-hand side of (11) cannot be attained by a semiclassical measurement. However, the first term of
(10) and the left-hand side of (13) can be asymptotically attained by a 1-semiclassical measurement, i.e., they can be asymptotically attained by measurements without using quantum correlations between samples. Thus, in order to attain it asymptotically, it is sufficient to physically realize the optimal measurement \( \Pi_1 \) on a single sample. Indeed, \( \Pi_1 \) is a random measurement as follows. To denote \( \Pi_1 \) as a random measurement, we will define the simple measurement \( E_g \) whose measurable space \( \mathcal{P}(\mathcal{H}) \). For an element \( g \in \text{SU}(k) \), the vectors \( \phi_1(g), \ldots, \phi_k(g) \) in \( \mathcal{H} \) are defined as \( (\phi_1(g) \cdots \phi_k(g)) = g \). The measurement \( E_g \) is defined as \( E_g (|\phi_i(g)\rangle \langle \phi_i(g)|) = |\phi_i(g)\rangle \langle \phi_i(g)| \). Therefore, the optimal measurement \( \Pi_1 \) for a single sample can be described as the following random measurement:

\[
\Pi_1 = \int_{\text{SU}(k)} E_g \mu(dg)
\]

where \( \mu \) is the invariant measure on \( \text{SU}(k) \) with \( \mu(\text{SU}(k)) = 1 \). Therefore, in order to realize the optimal measurement \( \Pi_1 \), it is sufficient to realize the simple measurement \( E_g \) for any \( g \in \text{SU}(k) \).

6. Conclusion

We have compared two cases. One regards the system consisting of enough samples as the single system, the other regards it as separate systems. Under this comparison, the error mean squares of both cases asymptotically coincide in the first order with respect to the Fubini-Study distance (see (10) and (19)). However, we leave the question of whether they asymptotically coincide in the second order with respect to the Fubini-Study distance to a future study. On the other hand, in view of the evaluation of large deviation, if the allowable radius is finite, neither coincide (see (11) and (20)). However, if the allowable radius goes to infinitesimal, both coincide (see (13) and (21)).

These results depend on the effect of a pure state. Therefore, it is an open question as to whether the error mean squares of both cases asymptotically coincide in the first order in another family. In the case of large deviation, the same question is also open in the limit where the radius of allowing error goes to infinitesimal.

Acknowledgments

The author wishes to thank Professor K. Ueno. Also, he wishes to thank Professor A. Fujiwara, Dr. K. Matsumoto and Professor H. Nagaoka.
References


CHAPTER 24
Optimal Universal Quantum Cloning and State Estimation

Dagmar Bruß, Artur Ekert, and Chiara Macchiavello

Abstract. We derive a tight upper bound for the fidelity of a universal $N \rightarrow M$ qubit cloner, valid for any $M \geq N$, where the output of the cloner is required to be supported on the symmetric subspace. Our proof is based on the concatenation of two cloners and the connection between quantum cloning and quantum state estimation. We generalize the operation of a quantum cloner to mixed and/or entangled input qubits described by a density matrix supported on the symmetric subspace of the constituent qubits. We also extend the validity of optimal state estimation methods to inputs of this kind.

Perfect quantum cloning is impossible [9]. This notwithstanding, we may ask how well we can clone quantum states. Bužek and Hillery, who were the first to address this problem, provided an example of a quantum device which can clone an unknown pure state of a single input qubit (a two-state system) into two output qubits, albeit with a certain fidelity smaller than one [2]. Their construction was subsequently shown to be optimal by Bruß et al. [1]. In this Letter we derive the optimal fidelity of a universal and symmetric quantum cloning machine (QCM) which acts on $N$ original qubits and generates $M$ clones.

A universal $N \rightarrow M$ quantum cloner is a quantum machine which performs a prescribed unitary transformation on an extended input which contains $N$ original qubits, $M - N$ “blank” qubits, and $K$ auxiliary qubits, and which outputs $M$ clones together with $K$ auxiliary qubits. The original qubits are all in the same (unknown and pure) quantum state described by the density operator $\varrho_{\text{in}} = \frac{1}{2}(1 + \vec{s}_{\text{in}} \cdot \vec{\sigma})$, where $\vec{s}_{\text{in}}$ is the original Bloch vector. Both “blanks” and the auxiliary qubits are initially in some prescribed quantum state. The output qubits are in an entangled state and in the present work we require that the density operator describing the state of the $M$ clones is supported on the symmetric subspace. This guarantees

that all the output qubits are indistinguishable and in the same state described by the reduced density operator $\varrho^{\text{out}}$. We comment on relaxing this assumption at the end of the paper.

It has been shown that universal $1 \to 2$ cloners can shrink only the original Bloch vector, without changing its orientation in the Bloch sphere [1]. The same argument as given in [1] (namely the impossibility to find a transformation that rotates any Bloch vector of the one-particle reduced density matrix by the same angle) applies also generally for an $N \to M$ qubit cloner. Therefore, the operation of a universal QCM can be characterized by the shrinking factor $\eta(N, M)$, which is also known in the literature as the Black Cow factor [8], and the reduced output density operator is of the form $\varrho^{\text{out}} = \frac{1}{2}[1 + \eta(N, M)s^{\text{in}} \cdot \vec{\sigma}]$. Universal $N \to M$ quantum cloning machines may be constructed in many different ways; the best constructions are those which maximize $\eta(N, M)$ (i.e., which maximize the fidelity of the cloning machine) and we refer to them as the optimal cloners.

Gisin and Massar have constructed a class of universal $N \to M$ QCMs and showed, using numerical methods, that for $N \leq 7$ their cloners are optimal [3]. Our derivation of the upper bound for $\eta(N, M)$ is quite general and does not refer to any specific realization of the universal cloning machines. In particular, it shows that the Gisin-Massar cloners saturate this bound for any $N$ and $M \geq N$. Our approach avoids an elaborate optimization procedure, extends the class of allowed inputs to mixed and/or entangled states of the original qubits which belong to the symmetric subspace, and sheds some light on the connection between optimal quantum cloning and optimal quantum state estimation. The proof is based on the concatenation of two quantum cloners and on associating the upper bound on the fidelity of an $M \to \infty$ cloner with the fidelity of the optimal state estimation of $M$ qubits, given in [5].

We concatenate two cloning machines in the following way. The first cloner is an $N \to M$ universal machine characterized by the shrinking factor $\eta(N, M)$. The $M$ clones from the output of the first cloner are then taken as originals for the input into the second cloning machine which creates infinitely many clones with the shrinking factor $\eta(M, \infty)$. We now write down two statements which will be proved after unfolding the main result: (a) The shrinking factors for concatenated cloners multiply. (b) The equality

$$\eta_{\text{QCM}}^{\text{opt}}(M, \infty) = \eta_{\text{meas}}^{\text{opt}}(M)$$

holds. Here $\eta_{\text{meas}}^{\text{opt}}(M)$ corresponds to the optimal state estimation derived
in [5], and its meaning will be explained below.

Because of statement (a), the shrinking factors of universal cloning machines in sequence multiply. Moreover, the sequence of the two machines cannot perform better than the optimal \(N \to \infty\) universal cloner, otherwise the \(N \to \infty\) universal cloner would not be optimal. Thus we arrive at the following inequality:

\[
\eta_{QCM}(N, M) \cdot \eta_{QCM}(M, \infty) \leq \eta_{QCM}^{\text{opt}}(N, \infty). \tag{2}
\]

This means that the lowest upper bound for the general \(N \to M\) cloner is given by

\[
\eta_{QCM}(N, M) \leq \frac{\eta_{QCM}^{\text{opt}}(N, \infty)}{\eta_{QCM}^{\text{opt}}(M, \infty)}. \tag{3}
\]

We have thus reduced the optimality problem of the \(N \to M\) cloner to the task of finding the optimal \(N \to \infty\) cloner.

Now we can use statement (b) and the explicit form of \(\bar{\eta}_{\text{meas}}^{\text{opt}}(M)\) (see [5]), namely

\[
\bar{\eta}_{\text{meas}}^{\text{opt}}(M) = \frac{M}{M + 2} \tag{4}
\]

to conclude the central result that for any \(M \geq N\)

\[
\eta_{QCM}^{\text{opt}}(N, M) = \frac{NM + 2}{MN + 2}. \tag{5}
\]

For pure input states this corresponds to the optimal fidelity

\[
F_{QCM}^{\text{opt}}(N, M) = \frac{NM + N + M}{M(N + 2)}, \tag{6}
\]

which is achieved by the cloning transformations proposed in [3]. (For \(\rho^{\text{in}} = |\psi\rangle\langle \psi|\) the fidelity is defined as \(F = \langle \psi|\rho^{\text{out}}|\psi\rangle\).)

Let us note in passing that as the consequence of the factorization property (3) we can produce \(M\) clones from \(N\) originals either by applying directly the optimal \(N \to M\) cloner or by taking any number of intermediate steps in order to realize the cloning process, using the optimal transformation at each step; both ways lead to the same overall shrinking factor.

Let us now justify statements (a) and (b).

In order to prove (a) we describe an \(N \to M\) cloner in terms of a completely positive map \(C_{NM}\) which maps input density operators of \(N\) identical pure originals into output density operators of \(M\) clones, such that for any state \(|\psi\rangle\langle \psi|\) of a single input qubit (original) we have

\[
\text{Tr}_{M-1}[C_{NM}(|\psi\rangle\langle \psi|^{\otimes n}) = \eta(N, M)|\psi\rangle\langle \psi| + [1 - \eta(N, M)]\frac{1}{2}1], \tag{7}
\]
where the trace is performed on any $M - 1$ qubits (for an overview of completely positive operators, see [4]).

Let $\varrho_N$ be a density operator of $N$ qubits which is supported on the symmetric subspace of the $2^N$ dimensional Hilbert space. We can always write $\varrho_N$ as a linear combination of direct products of identical pure states, $\varrho_N = \sum_i \alpha_i |\psi_i\rangle\langle\psi_i|^{\otimes N}$, where $\sum_i \alpha_i = 1$; N.B. we do not require that all values $\alpha_i$ are positive [6,8]. The linearity of the completely positive map and its universality, i.e., the fact that $\eta(N,M)$ does not depend on $|\psi\rangle$, allow us to extend Eq. (7) to the more general form,

$$\text{Tr}_{M-1}[C_{NM}(\varrho_N)] = \eta(N,M)\varrho + [1 - \eta(N,M)]\frac{1}{2\cdot2} \mathbf{1}, \quad (8)$$

where $\varrho = \text{Tr}_{N-1}[\varrho_N]$. Now, suppose that we concatenate an $N \rightarrow M$ and an $M \rightarrow L$ cloner and view it as an $N \rightarrow L$ cloner. It evolves the initial $N$ qubit state $\varrho_N$ first into the $M$ qubit state $\varrho'_M$ and then into the $L$ qubit state $\varrho''_L$. The corresponding single qubit reduced density operators are $\varrho$, $\varrho' = \text{Tr}_{M-1}[\varrho'_M]$ and $\varrho'' = \text{Tr}_{L-1}[\varrho''_L]$. Following Eq. (8) we can write

$$\varrho'' = \eta(M, L)\varrho' + [1 - \eta(M,L)]\frac{1}{2\cdot2} \mathbf{1}$$

$$= \eta(N, M) \cdot \eta(M, L)\varrho + [1 - \eta(N,M) \cdot \eta(M,L)]\frac{1}{2\cdot2} \mathbf{1}, \quad (9)$$

i.e., indeed $\eta(N, L) = \eta(N, M) \cdot \eta(M, L)$.

We will now prove statement (b). Equation (5) was obtained assuming the following result (due to [5]): given $M$ qubits all in an unknown quantum state $|\psi\rangle$ there exists a universal POVM measurement $\{P_\mu\}$ [10] which leads to the best possible estimation of $|\psi\rangle$ with fidelity $\overline{F}(M) = \frac{M + 1}{M + 2}$, or, equivalently, with $\overline{\eta}(M) = \frac{M}{M + 2}$. The outcome of each instance of the measurement provides, with probability $p_\mu(\psi) = \text{Tr}(P_\mu|\psi\rangle\langle\psi|^{\otimes M})$, the “candidate” $|\psi_\mu\rangle$ for $|\psi\rangle$. The fidelity $\overline{F}_{\text{meas}}(M)$ is then calculated from the outcomes of the measurement as

$$\overline{F}_{\text{meas}}(M) = \sum_\mu p_\mu(\psi)|\langle \psi | \psi_\mu \rangle|^2 = \langle \psi | \overline{\varrho} | \psi \rangle, \quad (10)$$

where $\overline{\varrho} = \sum_\mu p_\mu(\psi)|\psi_\mu\rangle\langle\psi_\mu|$. In the optimal universal state estimating procedure the fidelity must not depend on $\psi$, thus $\overline{\varrho}$ can also be written as

$$\overline{\varrho} = \overline{\eta}_{\text{meas}}^{\text{opt}}(M)|\psi\rangle\langle\psi| + [1 - \overline{\eta}_{\text{meas}}^{\text{opt}}(M)]\frac{1}{2\cdot2} \mathbf{1}. \quad (11)$$

The optimal measurement of this type can be viewed as an $M \rightarrow \infty$ cloner because after reading each outcome we can prepare any number of
candidates, in particular infinitely many of them, with the average reconstruction fidelity \( F_{\text{meas}}^{\text{opt}}(M) \) with respect to the originals. Clearly, this procedure cannot provide a larger shrinking factor than the optimal \( M \to L \) cloner and we find
\[
F_{\text{meas}}^{\text{opt}}(M) \leq \eta_{\text{QCM}}^{\text{opt}}(M, L) \tag{12}
\]
for any \( L \geq M \), in particular for \( L \to \infty \).

Let us now show that for \( L \to \infty \) the formula (12) becomes the equality. To see this let us concatenate an \( M \to L \) cloner with a subsequent optimal state estimating measurement. The input to the cloner is of the form \( |\psi\rangle \langle \psi|^{\otimes M} \) and the output is described by the density operator \( \rho_L \) which is of the form
\[
\sum_i \alpha_i |\psi_i\rangle \langle \psi_i|^{\otimes L},
\]
where \( \sum_i \alpha_i = 1 \). The reduced density operator of each output qubit is
\[
\rho_i = \text{Tr}_L \left( |\psi_i\rangle \langle \psi_i|^{\otimes L} \right) = \eta(M, L) |\psi_i\rangle \langle \psi_i| + [1 - \eta(M, L)] \frac{1}{2} 1.
\]
The cloner \( M \to L \) concatenated with the state estimation on \( L \) qubits can be viewed as the state estimation performed on \( M \) qubits. The total procedure gives the fidelity of estimating \( |\psi\rangle \) which can be written as
\[
F_{\text{meas}}(M) = \langle \psi | \sum_{\mu} \text{Tr}(P_{\mu, L}) |\psi_\mu\rangle \langle \psi_\mu| \psi \rangle
\]
\[
= \sum_{\mu, i} \langle \psi | \alpha_i \text{Tr}(P_{\mu} |\psi_i\rangle \langle \psi_i|^{\otimes L}) |\psi_\mu\rangle \langle \psi_\mu| \psi \rangle \tag{13}
\]
\[
= \sum_i \langle \psi | \alpha_i \left[ F_{\text{meas}}^{\text{opt}}(L) |\psi_i\rangle \langle \psi_i| + [1 - F_{\text{meas}}^{\text{opt}}(L)] \frac{1}{2} 1 \right] |\psi \rangle \tag{14}
\]
which for \( L \to \infty \) becomes [due to \( F_{\text{meas}}^{\text{opt}}(L) \to 1 ]
\[
F_{\text{meas}}(M) \to \sum_i \langle \psi | \alpha_i |\psi_i\rangle \langle \psi_i| \psi \rangle = \langle \psi | \rho |\psi \rangle
\]
\[
= \frac{1}{2} [1 + \eta_{\text{QCM}}(M, \infty)]. \tag{15}
\]

The concatenation of a cloner with a measurement cannot perform better than the optimal measurement, thus we can write
\[
\eta_{\text{QCM}}^{\text{opt}}(M, \infty) \leq F_{\text{meas}}^{\text{opt}}(M) \tag{16}
\]
Combining Eqs. (12) and (16) finally leads to
\[
\eta_{\text{QCM}}^{\text{opt}}(M, \infty) = F_{\text{meas}}^{\text{opt}}(M), \tag{17}
\]
thus proving statement (b).

Before concluding, we want to stress that, as a consequence of what was shown above, we can extend the operation of a cloning machine and the
application of an optimal measurement to any input density operator of \(N\) qubits which has support on the symmetric subspace. The properties of the universal cloning machine as defined at the beginning of this paper allow us to conclude that the same machine can operate any such symmetric density operator and shrinks the Bloch vector of the reduced input density matrix by a fixed amount, independent of the initial length. Notice also that the optimal machine, for products of pure inputs specified by the shrinking factor (5), is still optimal for this extended class of inputs. Actually, if a better cloning machine existed for mixed input states, we would use it as the second cloner \(M \to \infty\) in Eq. (2), giving a smaller lower bound in Eq. (3). This would lead to a contradiction because we already know that universal cloners for pure states saturating the bound (5) exist [3].

One may want to relax our restriction and consider quantum cloners which produce identical clones (i.e., with the same single-qubit density operator \(\varrho^\text{out}\)), but for which the state of all outputs does not belong to the symmetric subspace. This case, in principle, may provide a higher shrinking factor; however, we could neither prove nor disprove this with our approach. We leave this problem as a challenge to other colleagues.

In a similar way as for the cloner we can extended the validity of an optimal universal measurement procedure to inputs from the symmetric subspace. In this case the goal is to find the optimal measurement for the reduced density operator for each input copy. Since we require the process to be universal, we know that the reduced density operator reconstructed as the result of the measurement given in Eq. (11) is just the shrunk version of the initial one. We can then describe the quality of the procedure in terms of the shrinking factor. We conclude that the optimal measurement derived in Ref. [5] is also optimal for any input symmetric state. Actually, if this were not the case, we could devise a measurement procedure on \(N\) initial pure qubits by first applying an \(N \to M\) cloner and then an optimal measurement on the mixed state of the output \(M\) clones. If this global measurement were better than the optimal one of Ref. [5] we would then obtain with the above procedure a universal measurement for pure states which performs better than the one in Ref. [5], thus finding a contradiction.

Let us mention in passing that in our discussion we found it convenient to use the shrinking factor, because it has an intuitive geometrical meaning both for pure and mixed states; however, one can rephrase the optimality argument for universal operations using, for example, the Uhlmann fidelity [7] for the reduced density operators.

In conclusion, we have derived the optimal shrinking factor/fidelity for
a universal $N \to M$ cloner and generalized its operation to a more general case of mixed and/or entangled input states which belong to the symmetric subspace. Furthermore, we have established the connection between optimal quantum state estimation and optimal quantum cloning which allowed us to extend the validity of the optimal state estimation methods [5] to inputs of the above form.

Acknowledgments

We would like to express our special thanks to J.I. Cirac, who helped us to clarify the ideas presented in the paper. It is a pleasure to thank C.H. Bennett, V. Bužek, D. DiVincenzo, N. Gisin, M. Palma, S. Popescu, and R. Werner for helpful discussions. This work was supported in part by the European TMR Research Network ERP-4061PL95-1412, Hewlett-Packard, The Royal Society of London, and Elsag-Bailey, a Finmeccanica Company.

References

CHAPTER 25

Bounds for Generalized Uncertainty of Shift Parameter

Alexander S. Holevo

1. Introduction

Let $\Omega$ be a Borel space with a measure $\mu$ and $\{p_\theta\}$ be a family of probability densities on $(\Omega, \mu)$, depending on the real parameter $\theta$. Under the assumption that the function $\theta \to \sqrt{p_\theta}$ with values in $L^2(\Omega, \mu)$ is strongly differentiable, the Cramér-Rao inequality for any complex random variable $Y$ on $(\Omega, \mu)$ with finite second moment $E_\theta \{ |Y|^2 \}$ holds:

$$
E_\theta \left\{ \left| Y - E_\theta \{ Y \} \right|^2 \right\} \geq \left| \frac{d}{d\theta} E_\theta \{ Y \} \right|^2 \cdot J_\theta^{-1},
$$

where $J_\theta = 4 \int_\Omega \left( \frac{d}{d\theta} \sqrt{p_\theta(w)} \right)^2 \mu(d\omega)$ is the Fisher information, supposed to be non-zero (see, e.g. [3]). If $X$ is arbitrary real random variable on $(\Omega, \mu)$, then substituting $Y = \exp(i\lambda X)$ in (1), we obtain an inequality for the characteristic function $\varphi_\theta(\lambda) = E_\theta \{ \exp(i\lambda X) \}$ of the random variable $X$:

$$
1 - |\varphi_\theta(\lambda)|^2 \geq \left| \frac{d}{d\theta} \varphi_\theta(\lambda) \right|^2 \cdot J_\theta^{-1},
$$

(2)

Assume now that the family $\{p_\theta\}$ is invariant under arbitrary shifts of the parameter $\theta$, i.e., under the action of the additive group of reals $\mathbb{R}$. Then, if $X$ is an invariant estimate of the parameter $\theta$, its characteristic function satisfies the relation

$$
\varphi_{\theta+\Delta}(\lambda) = \exp(i\lambda \Delta) \varphi_\theta(\lambda), \quad \Delta \in \mathbb{R},
$$

(3)

whence

$$
\frac{d}{d\theta} \varphi_\theta(\lambda) = i\lambda \varphi_\theta(\lambda).
$$

(4)

Editorial note: The ideas of this paper are further developed in Holevo [IV-4]. See the references of Chap. 21.
Put
\[\Delta(\lambda) = \lambda^{-2} \left\{ |\varphi_\theta(\lambda)|^{-2} - 1 \right\}, \quad \lambda \neq 0. \quad (5)\]

Then (3) implies that $\Delta(\lambda)$ does not depend on $\theta$. Apparently $0 \leq \Delta(\lambda) \leq +\infty$. From (2) we deduce the inequality
\[\Delta(\lambda) \geq J^{-1}_\theta, \quad \lambda \neq 0. \quad (6)\]

It follows that $\Delta(\lambda), \lambda \neq 0$, can be regarded as a functional measure of uncertainty of the invariant estimate $X$. If the variance $D_\theta\{X\}$ of the estimate is finite, then $D_\theta\{X\} = \lim_{\lambda \to 0} \Delta(\lambda)$, and (6) implies the ordinary Cramér-Rao inequality; however the quantity $\Delta(\lambda)$ may be finite for estimates with infinite variance.

The relation (6) has a curious corollary (which, however, will not be used in the sequel). Let $p(x)$ be a probability density on $\mathbb{R}$; put $p_\theta(x) = p(x - \theta)$ and assume that $p_\theta(x)$ satisfies the condition formulated at the beginning of this section. Then (6) implies the inequality for the characteristic function $\varphi(\lambda)$ of the density $p(x)$
\[|\varphi(\lambda)|^2 \leq \left( 1 + \lambda^2 \cdot J^{-1} \right)^{-1}, \quad (7)\]
where $J = 4 \int_{\mathbb{R}} \left[ \sqrt{p(x)} \right]^2 dx$. Integrating and using the Parseval relation we also obtain $\int p(x)^2 dx \leq \sqrt{J}$. It is not difficult to generalize all these considerations to the case $\theta \in \mathbb{R}^n$.

Let now $\theta$ be an angular parameter, then the symmetry group is the additive (mod $2\pi$) group of the segment $[0, 2\pi]$, which will be denoted $\mathbb{T}$. Then the relations (2)–(7) hold only for all integer $\lambda$, i.e., for $\lambda$ belonging to the group $\mathbb{Z}$ with is the dual of $\mathbb{T}$. Putting $\lambda = 1$ in (5), (6) we obtain an analog of the Cramér-Rao inequality for the angular parameter
\[|E_\theta \{\exp(iX)\}|^{-2} - 1 \geq J^{-1}_\theta \quad (8)\]

These simple results can be regarded as a motivation for their “noncommutative” analogs, which will be given below. In quantum theory one meets well-known difficulties if one tries to extends the Heisenberg uncertainty relation to canonical pairs of observables, different from the “position-momentum” pairs (see [1, 4] for discussion). The inequality of Section 2 gives a uniform quantitative expression of the uncertainty principle for any canonical pair of interest. In Section 3 we give a better, attainable, lower bound for the generalized uncertainty of a shift parameter of quantum state and show that this bound is asymptotically equivalent to the bound of the uncertainty principle in the case of $n$ independent observations, when $n \to \infty$. 
2. Quantum Covariant Family

Let \( \{ S_\theta; \theta \in \Theta \} \) be a family of quantum states (density operators) in a Hilbert space \( H \). We shall consider families \( \{ S_\theta \} \), invariant under shifts of the parameter \( \theta \). This means that either \( \Theta = \mathbb{R} \) or \( \Theta = \mathbb{T} \) and

\[
S_\theta = e^{-i\theta A} S e^{i\theta A}, \quad \theta \in \Theta
\]

where \( \theta \to e^{-i\theta A} \) is a unitary representation of the group \( \Theta \) in the space \( H \). Correspondingly, \( \lambda \) will run over the dual group \( \hat{\Theta} = \mathbb{R} \) or \( \hat{\Theta} = \mathbb{Z} \).

Assume that the self-adjoint operator \( A \) has finite second moment with respect to the states \( S \). Then for any bounded operator \( Y \) the following inequality holds:

\[
E_\theta \left\{ (Y - E_\theta \{ Y \})^* \circ (Y - E_\theta \{ Y \}) \right\} \geq \left| \frac{d}{d\theta} E_\theta \{ Y \} \right|^2 \cdot [4D_\theta \{ A \}]^{-1},
\]  

where \( X \circ Y = \frac{1}{2}(XY + YX) \), \( E_\theta \{ X \} = \operatorname{Tr} S_\theta X \), and \( D_\theta \{ X \} \) is expressed through \( E_\theta \{ X \} \) by the usual formula \[4\]. One obtains stronger inequality if \( 4D_\theta \{ A \} \) is replaced by the noncommutative analog of the Fisher information; the resulting inequality is a strict analog of (1), but we shall not need it.

Let \( M = \{ M(du) \}, u \in \Theta \), be a quantum measurement of the parameter \( \theta \), i.e., a resolution of identity in the Hilbert space \( H \) \[4\]. The probability distribution of the measurement \( M \) with respect to the state \( S_\theta \) is given by the relation \( \mu^M_\theta(du) = \operatorname{Tr} S_\theta M(du) \). Denoting by \( \phi^M_\theta(\lambda) \) the characteristic function of the distribution \( \mu^M_\theta \), we have \( \phi^M_\theta(\lambda) = \operatorname{Tr} S_\theta U(\lambda) \), where

\[
U(\lambda) = \int_{\Theta} \exp(i\lambda \theta) M(d\theta).
\]  

(10)

Note that \( ||U(\lambda)|| \leq 1 \). Substituting \( Y = U(\lambda) \) into (9) and taking into account that \( U(\lambda)^* \circ U(\lambda) \leq I \), we obtain

\[
1 - |\phi^M_\theta(\lambda)|^2 \geq \left| \frac{d}{d\theta} \phi^M_\theta(\lambda) \right|^2 \cdot [4D_\theta \{ A \}]^{-1}
\]  

(11)

Let now \( M \) be a covariant measurement of the parameter \( \theta \), so that \( \mu^M_{\theta + \Delta}(B) = \mu^M_\theta(B - \Delta) \) for all \( \Delta \in \Theta \), where in the case \( \Theta = \mathbb{T} \) we mean the addition (mod \( 2\pi \)). Then \( \phi^M_\theta(\lambda) \) satisfies (3), (4). Denoting by \( \Delta^M_S(\lambda) \) the quantity defined through \( \phi_\theta(\lambda) = \phi^M_\theta(\lambda) \) by the relation (5), i.e., \( \Delta^M_S(\lambda) = \lambda^{-2} \cdot \left\{ \int \exp(i\lambda \theta) \mu^M_S(d\theta) \right\}^{-2} - 1 \) we obtain as in (6)

\[
\Delta^M_S(\lambda) \cdot D_S \{ A \} \geq \frac{1}{4},
\]  

(12)
where we have put \( D_S\{A\} \equiv D_\theta\{A\} \), since this is independent of \( \theta \).

In quantum theory \( \theta \) is a kinematical parameter such as position, time, angle, phase, and the operator \( A \) corresponds to quantities which would be canonically conjugate to \( \theta \) in classical mechanics - momentum, energy, angular momentum, action. A covariant measurement of the kinematical parameter \( \theta \) determines a quantum-mechanical quantity which is in natural sense canonically conjugate to the observable \( A \). The inequality (12) gives then a general quantitative expression of the uncertainty principle for quantum canonically conjugate quantities. If \( \theta \in \mathbb{R} \) as it is in the cases of position and time, then (12) implies the uncertainty relation for variances of the canonically conjugate quantities. If \( \theta \in \mathbb{R} \) as in the cases of angle and phase, then one may take \( \lambda = 1 \) to obtain the inequality of type (8) for the angular uncertainty.

3. Pure State Case

Let \( S \) be a pure state. In this case it is possible to find explicitly the value of \( \Delta^*_S(\lambda) = \min_M \Delta^M_S(\lambda) \), where the minimum is taken over all covariant measurements of the parameter \( \theta \). It turns out that for all \( \lambda \) the minimum is attained for one and the same optimal measurement \( M^* \). The quantity \( \Delta^*_S(\lambda) \) may be interpreted as an “internal (independent of the particular measurement) uncertainty” of the parameter \( \theta \) in the state \( S \).

For definiteness we consider the case \( \Theta = \hat{\Theta} = \mathbb{R} \). Assume that the operator \( A \) has purely Lebesgue spectrum (this is the necessary and sufficient condition for the existence of a covariant measurement of a parameter \( \theta \in \mathbb{R} \)). Then

\[
H = \int \oplus H(\lambda) d\lambda,
\]

so that \( \exp(-i\theta A)\varphi = \int \oplus \exp(-i\theta \lambda)\varphi(\lambda) d\lambda \), if \( \varphi = \int \oplus \varphi(\lambda) d\lambda \) [2]. Let \( S = |\psi\rangle\langle \psi| \), where \( \psi = \int \oplus \psi(\lambda) d\lambda \) is a unit vector of \( H \), so that \( \int \|\psi(\lambda)\|^2 = 1 \). The function \( \|\psi(\lambda)\|^2 \) has the meaning of probability density of the observable \( A \) in the state \( S \), and the finiteness of the second moment of \( A \) means that

\[
\int \lambda^2 \|\psi(\lambda)\|^2 d\lambda < \infty. \tag{13}
\]

In [4] section IV.11, it is shown that for any covariant measurement \( M \) the characteristic function \( \varphi^M_\theta(\lambda) \) satisfies

\[
|\varphi^M_\theta(\lambda)| \leq \varphi^*(\lambda) = \int \|\psi(\mu)\| \cdot \|\psi(\mu + \lambda)\| d\mu. \tag{14}
\]
The sign of equality here is achieved for all \( \lambda \) by some \( M^* \) independent of \( \lambda \). We shall not write down the explicit formula for \( M^* \) since it will not be needed. From (14) we have

\[
\Delta_S^M(\lambda) \geq \Delta^*(\lambda) = \lambda^{-2}[\varphi^*(\lambda)^2 - 1].
\]  

(15)

Consider the quantum-mechanical analog of the probabilistic scheme of \( n \) independent observations. Let \( H = H^{(1)} \otimes \cdots \otimes H^{(1)} \), \( \psi = \psi^{(n)} = \psi^{(1)} \otimes \cdots \otimes \psi^{(1)} \), where \( \psi^{(1)} \) is a unit vector in \( H^{(1)} \), so that \( S = S^{(1)} \otimes \cdots \otimes S^{(1)} \), where \( S^{(1)} = |\psi^{(1)}\rangle \langle \psi^{(1)}| \), and let \( e^{-i\theta A} = e^{-i\theta A^{(1)} \otimes \cdots \otimes e^{-i\theta A^{(1)}}} \), where \( A^{(1)} \) is an operator in \( H^{(1)} \). Then \( D_S \{A\} = nD_{S^{(1)}} \{A^{(1)}\} \), and the inequality (12) reads

\[
\Delta_S^M(\lambda) \geq [4nD_{S^{(1)}} \{A^{(1)}\}]^{-1}.
\]  

(16)

We shall show that under some regularity assumptions the bound (16) is asymptotically (with \( n \to \infty \)) equivalent to the bound (15). Thus the generalized uncertainty principle of Section 2 gives the bound which is asymptotically attainable by the covariant measurement \( M^* \).

We have

\[
\psi^{(n)}(\lambda) = \int \cdots \int \psi^{(1)}(\lambda - \lambda_2 - \cdots - \lambda_n) \otimes \psi^{(1)}(\lambda_2) \otimes \cdots \otimes \psi^{(1)}(\lambda_n) d\lambda_2 \cdots d\lambda_n,
\]

from which it follows that \( \|\psi^{(n)}(\lambda)\|^2 \) is the \( n \)-fold convolution of the function \( \|\psi^{(1)}(\lambda)\|^2 \) with itself. Thus \( \|\psi^{(n)}(\lambda)\|^2 \) is the probability density of the sum of \( n \) independent random variables with the density \( \|\psi^{(1)}(\lambda)\|^2 \), satisfying (13). Put

\[
m = E_{S^{(1)}} \{A^{(1)}\} = \int \lambda \|\psi^{(1)}(\lambda)\|^2 d\lambda,
\]

\[
\sigma^2 = D_{S^{(1)}} \{A^{(1)}\} = \int (\lambda - m)^2 \|\psi^{(1)}(\lambda)\|^2 d\lambda,
\]

and consider the densities \( p^{(n)}(\lambda) = \|\psi^{(n)}(\lambda)\|^2 \) of the corresponding normalized random variables. Then the result of [5] implies that \( p^{(n)}(\lambda) \to p(\lambda) \) in \( L^1(\mathbb{R}) \), where \( p(\lambda) \) is the density of standard normal distribution. Taking into account the inequality

\[
\int \left| \sqrt{f(\lambda)} - \sqrt{g(\lambda)} \right|^2 d\lambda \leq \int |f(\lambda) - g(\lambda)| d\lambda,
\]  

(17)

we obtain that \( \sqrt{p^{(n)}(\lambda)} \to \sqrt{p(\lambda)} \) in \( L^2(\mathbb{R}) \). Hence

\[
\varphi^*(\lambda \sqrt{n} \sigma) = \int \sqrt{p^{(n)}(\mu)} \sqrt{p^{(n)}(\mu + \lambda)} d\mu
\]

\[
\to \int \sqrt{p(\mu)} \cdot \sqrt{p(\mu + \lambda)} d\mu = \exp(-\lambda^2/8),
\]
from which it follows that the probability distribution of the optimal measurement $M^*$ with respect to the state $S$ is asymptotically normal with parameters $(0, [4n\sigma^2]^{-1})$ (Since $M^*$ is covariant, it follows that its probability distribution with respect to the state $S_\theta$ is asymptotically normal with parameters $(\theta, [4n\sigma^2]^{-1})$).

Yet this does not imply that the bound (16) is asymptotically attainable, for one must know the asymptotics of the quantity

$$\varphi^*(\lambda) = \int \sqrt{p^{(n)}(\mu)} \cdot \sqrt{p^{(n)}(\mu + \lambda/\sqrt{n\sigma})} \, d\mu$$

$$= 1 - \frac{1}{2} \int \left[ \sqrt{p^{(n)}(\mu + \lambda/\sqrt{n\sigma})} - \sqrt{p^{(n)}(\mu)} \right]^2 \, d\mu,$$  \hspace{1cm} (18)

entering in the right-hand side of (15), with $\lambda$ fixed. For this we use the asymptotic expansion of $p^{(n)}(\lambda)$ in $L^1$ metric, established in [5]. Assume that $\int |\lambda|^5 \|\psi^{(1)}(\lambda)\|^2 \, d\lambda = MS\{|A|^5\} < \infty$. Then there are polynomials $P^{(1)}(\lambda), P^{(2)}(\lambda)$ such that

$$\int \left| p^{(n)}(\lambda) - p^{(n)}(\lambda) \left[ 1 + \frac{1}{\sqrt{n}} P^1(\lambda) + \frac{1}{n} P^2(\lambda) \right] \right| \, d\lambda = o \left( \frac{\sqrt{n}}{n^{3/2}} \right) = o \left( \frac{1}{n} \right).$$

Taking into account (17), this implies

$$\sqrt{p^{(n)}(\lambda)} = \sqrt{p(\lambda)} \cdot \sqrt{1 + \frac{1}{\sqrt{n}} P^1(\lambda) + \frac{1}{n} P^2(\lambda) + \Delta_n(\lambda)},$$  \hspace{1cm} (19)

where $\int \Delta_n(\lambda)^2 \, d\lambda = o \left( \frac{1}{n} \right)$. From (18) and (19) after some estimates we obtain

$$\varphi^*(\lambda) = 1 - \frac{1}{2} \int \left[ \sqrt{p(\mu + \lambda/\sqrt{n\sigma})} - \sqrt{p(\mu)} \right]^2 \, d\mu + o \left( \frac{1}{n} \right)$$

$$= 1 - \lambda^2 / 8n\sigma^2 + o \left( \frac{1}{n} \right).$$

Substituting this into (15) we find $\Delta^*(\lambda) \sim [4n\sigma^2]^{-1}$, which means that (16) asymptotically attainable.

It seems that the condition of finiteness of the fifth moment of $A$ may be considerably relaxed; in the case $\Theta = \mathbb{T}$ the variable $\lambda$ is integer and one should use the corresponding local limit theorems for lattice random variables.
References


PART V

Large Deviation Theory in Quantum Estimation

Chap. 27: H. Nagaoka
“On the relation between Kullback divergence and Fisher information:
From classical systems to quantum systems” . . . . . . . . . . . . . . . . . . . . . . . . 399

Chap. 28: M. Hayashi
“Two quantum analogues of Fisher information from a large
deviation viewpoint of quantum estimation” . . . . . . . . . . . . . . . . . . . . . . . . 420

Chap. 29: M. Keyl and R.F. Werner
“Estimating the spectrum of a density operator” . . . . . . . . . . . . . . . . . . . . . . . 458
CHAPTER 26

Introduction to Part V

1. Large Deviation Analysis in Classical Estimation

In the preceding chapters, we mainly focused on mean square error (MSE) as a criterion for the estimation error. However, this is not necessarily the most suitable criterion in every case. Suppose we are interested only in the probability that the difference between the estimated parameter and the true one is larger than a given constant $\varepsilon > 0$. We also assume that the number $n$ of our observations is sufficiently large. In this case, we can focus on this probability instead of MSE. Typically, if we choose a suitable estimator, this probability goes to 0 exponentially as $n$ goes to infinity. Thus, it is natural to optimize the exponential rate of this probability. The theory concerning this exponential rate is called large deviation theory in estimation theory, and its classical version was initiated by Bahadur [V-1].

Here, we give its sketch for a one-parameter family in the classical setting. Suppose we have the statistical model of the $n$-fold tensor product of the model $\{p_\theta | \theta \in \mathbb{R}\}$. As reviewed in section 2 of chapter 27, if the given constant $\varepsilon$ is small enough, the optimal decreasing exponent approximately equals $\frac{J_\theta}{2} \varepsilon^2$. It can be attained by the maximum likelihood estimator in the asymptotic sense under some regularity conditions. The detail of this statement is the following: Assume that $T^{(n)}$ denotes the estimator based on $n$ i.i.d observations. We usually consider the weak consistency condition:

$$p^n_\theta \left\{ |T^{(n)} - \theta| \geq \varepsilon \right\} \rightarrow 0, \quad \forall \varepsilon > 0, \quad \forall \theta \in \mathbb{R}.$$  

If we denote the set of weakly consistent estimator sequences by $\mathcal{C}$, the following relations hold:

$$\limsup_{\varepsilon \downarrow 0} \sup_{\{T^{(n)}\} \in \mathcal{C}} \limsup_{n \to \infty} \frac{-1}{n \epsilon^2} \log p^n_\theta \left\{ |T^{(n)} - \theta| \geq \varepsilon \right\} = \sup_{\{T^{(n)}\} \in \mathcal{C}} \limsup_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{-1}{n \epsilon^2} \log p^n_\theta \left\{ |T^{(n)} - \theta| \geq \varepsilon \right\} = \frac{1}{2} J_\theta. \quad (1)$$

That is, the asymptotically optimal performance according to the large deviation criterion is characterized by Fisher information as in the MSE criterion.

395
2. First Large Deviation Analysis on Quantum Estimation

Nagaoka [Chap. 27] considered quantum analogues of (1). In the quantum state estimation with a quantum state family \( \{ \rho_\theta \in \mathbb{R} \} \), an estimator in the \( n \)-i.i.d. setting is described by a pair consisting of a quantum measurement \( M^n \) on the composite system \( \mathcal{H} \otimes n \) and a function \( T^{(n)} \) from the obtained data to the parameter space. He focused on a sequence of estimators \( \{(M^n, T^{(n)})\} \), and proved that

\[
\sup_{\{(M^n, T^{(n)})\} \in C} \limsup_{\varepsilon \downarrow 0} \limsup_{n \to \infty} -\frac{1}{n\varepsilon^2} \log P_{\rho_\theta}^{M^n} \{T^{(n)} \geq \theta + \varepsilon\} = \frac{1}{2} J_\theta \quad (2)
\]

\[
\limsup_{\varepsilon \downarrow 0} \sup_{\{(M^n, T^{(n)})\} \in C} \limsup_{n \to \infty} -\frac{1}{n\varepsilon^2} \log P_{\rho_\theta}^{M^n} \{T^{(n)} \geq \theta + \varepsilon\} = \frac{1}{2} \tilde{J}_\theta \quad (3)
\]

with an artificial definition of a class \( C \) of sequences of estimators, where \( J_\theta \) and \( \tilde{J}_\theta \) are called SLD and Kubo-Mori-Bogoljubov (KMB) Fisher information, respectively, and are defined as

\[
J_\theta \equiv \text{Tr} \rho_\theta L_\theta^2 = \text{Tr} \frac{d\rho_\theta}{d\theta} L_\theta, \quad \frac{1}{2} (L_\theta \rho_\theta + \rho_\theta L_\theta) = \frac{d\rho_\theta}{d\theta}
\]

\[
\tilde{J}_\theta \equiv \text{Tr} \int_0^1 \rho_\theta^\lambda \tilde{L}_\theta \rho_\theta^{1-\lambda} \tilde{L}_\theta d\lambda = \text{Tr} \frac{d\rho_\theta}{d\theta} \tilde{L}_\theta, \quad \tilde{L}_\theta = \frac{d\log \rho_\theta}{d\theta}.
\]

This result is considered remarkable in the following sense. Before this result was given, the KMB Fisher information had not been discussed sufficiently. A little discussion had been done only in the context of the geometry of quantum states while the SLD Fisher information had been studied mainly from the viewpoint of quantum estimation. Thus, equations (2) and (3) can be regarded as the first comparison between these two quantum analogues of Fisher information from a unified viewpoint.

Nagaoka also pointed out the necessity to investigate the following:

- Whether these equations hold when we put \( C \) as the set of weakly consistent estimators \( \{(M^n, T^{(n)})\} \), which is defined as sequences of estimators satisfying

\[
P_{\rho_\theta}^{M^n} \{|T^{(n)} - \theta| \geq \varepsilon\} \to 0, \quad \forall \varepsilon > 0, \quad \forall \theta \in \mathbb{R}.
\]

- Otherwise, what type of definition of \( C \) is suitable for the validity of these equations.

Moreover, he proved the following equation (4) as an alternative to (2),
but he did not prove the equation (5) to be corresponding to (3):

\[
\sup_{\{ (M^n, T^{(n)}) \} \in C} \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{-1}{n \varepsilon^2} \log P_{\rho_{\theta}^n}^n \{ |T^{(n)} - \theta| \geq \varepsilon \} = \frac{1}{2} \tilde{J}_\theta \tag{4}
\]

\[
\lim_{\varepsilon \downarrow 0} \sup_{\{ (M^n, T^{(n)}) \} \in C} \lim_{n \to \infty} \frac{-1}{n \varepsilon^2} \log P_{\rho_{\theta}^n}^n \{ |T^{(n)} - \theta| \geq \varepsilon \} = \frac{1}{2} \tilde{J}_\theta. \tag{5}
\]

3. Further Large Deviation Analysis on Quantum Estimation

For an indepth analysis of this topic, Hayashi [Chap. 28] introduced strong consistency conditions for sequences of estimators:

(S.1) The following convergence is uniform in the parameter \( \theta \):

\[
\alpha(\{ (M^n, T^{(n)}) \}, \theta) \equiv \liminf_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{-1}{n \varepsilon^2} \log P_{\rho_{\theta}^n}^n \{ |T^{(n)} - \theta| \geq \varepsilon \}.
\]

(S.2) The limit \( \alpha(\{ (M^n, T^{(n)}) \}, \theta) \) is continuous in \( \theta \).

He proved the following as alternatives to (2), (3), (4), and (5):

\[
\sup_{\{ (M^n, T^{(n)}) \} : SC} \liminf_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{-1}{n \varepsilon^2} \log P_{\rho_{\theta}^n}^n \{ |T^{(n)} - \theta| \geq \varepsilon \} = \frac{1}{2} \tilde{J}_\theta \tag{6}
\]

\[
\lim_{\varepsilon \downarrow 0} \sup_{\{ (M^n, T^{(n)}) \} : SC} \limsup_{n \to \infty} \frac{-1}{n \varepsilon^2} \log P_{\rho_{\theta}^n}^n \{ |T^{(n)} - \theta| \geq \varepsilon \} = \frac{1}{2} \tilde{J}_\theta. \tag{7}
\]

Furthermore, by constructing a weakly consistent sequence of estimators \( \{ (M^n, T^{(n)}) \} \) achieving the \( \frac{1}{2} \tilde{J}_\theta \) bound, he showed that the equations (2) and (4) do not hold when \( C \) is the set \( C_{WC} \) of weakly consistent estimators, and obtained the following equation:

\[
\sup_{\{ (M^n, T^{(n)}) \} \in C_{WC}} \liminf_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{-1}{n \varepsilon^2} \log P_{\rho_{\theta}^n}^n \{ |T^{(n)} - \theta| \geq \varepsilon \} = \frac{1}{2} \tilde{J}_\theta.
\]

Such a weakly consistent sequence of estimators can be regarded as a superefficient estimator from a mathematical statistical viewpoint.

4. Estimation of Eigenvalues

Next, we proceed to the estimation of the eigenvalues of the unknown density matrix. The first attempt at this problem has been done by Jozsa and the Horodecki family [V-2].

They treated the estimation problem of the entropy of the unknown density matrix in connection with the problem of fixed-length universal
quantum information compression. It can be regarded as the estimation of a functional of the eigenvalues of the unknown density matrix. They constructed their projections estimating the entropy by modifying the method of type*, and evaluated the exponent of the one-side error probability.

After this research, the estimation of the eigenvalues was independently discussed by Keyl & Werner [Chap. 29] and Matsumoto†. They proposed the same estimator using the representation theory on the tensor product space. They also obtained the same exponent of the error probability, which is characterized by the relative entropy $D(\hat{\rho}||\rho)$ between the true eigenvalues $p = (p_i)$ and the estimated eigenvalues $\hat{\rho} = (\hat{p}_i)$.

Based on these results, Hayashi and Matsumoto applied the estimation of the eigenvalues to universal concentration [V-4], and variable-length universal quantum information compression [V-3]. Hayashi [V-5] also applied it to a more detailed analysis of fixed-length universal quantum information.

Further Reading


*The method of type is a famous method in information theory.
†Matsumoto’s results was obtained by K. Matsumoto, seminar note, (1999), (In Japanese), for the first time. But, the essence of this part later appeared in [V-4] and [V-3].
CHAPTER 27

On the Relation between Kullback Divergence and Fisher Information: From Classical Systems to Quantum Systems

Hiroshi Nagaoka

Abstract. Kullback divergence $D(p \parallel q)$ for a couple of probability distributions $\{p, q\}$ and Fisher information $J(\theta)$ for a parametric family of probability distributions $\{p_\theta\}$ are very important notions in the classical theory of statistical inference. Quantum analogues of these notions are known as Umegaki’s relative entropy $D(\rho \parallel \sigma)$ for a couple of density operators $\{\rho, \sigma\}$ and the quantum Fisher information $J(\theta)$ based on the symmetric logarithmic derivatives for a parametric family of density operators $\{\rho_\theta\}$. In the classical case, the relation $D(p_{\theta+\Delta\theta} \parallel p_\theta) = \frac{1}{2} J(\theta) (\Delta \theta)^2 + o((\Delta \theta)^2)$ holds, which connects the asymptotic theory of hypothesis testing and that of parameter estimation. In the quantum case, the relation $D(\rho_{\theta+\Delta\theta} \parallel \rho_\theta) = \frac{1}{2} \tilde{J}(\theta) (\Delta \theta)^2 + o((\Delta \theta)^2)$ defines $\tilde{J}(\theta)$, which is not equal to $J(\theta)$ in general and is another quantum analogue of Fisher information. We show the inequality $\tilde{J}(\theta) \geq J(\theta)$ and elucidate its meaning in connection with the statistical inference for quantum states.

1. Introduction

The concept of probability plays a fundamental role in quantum mechanics. In this sense, it can be said that quantum mechanics is based on probability theory. At the same time, however, quantum mechanics can be regarded as a variant or a generalization of probability theory. (This standpoint is formulated in the most natural way by using a language of operator algebras, although we do not treat it here; see e.g. [15].) For instance, there are lots of analogies between those concepts as quantum states (density operators) and probability distributions (probability measures), quantum mechanical observables and random variables, etc., and through these correspondences we can study quantum mechanics and probability theory in a unified manner. In this paper, we investigate a fundamental problem concerning statistical inference for probability distributions and its quantum

This chapter was originally written in Japanese. It was translated to English by the author.
version. (We sometimes use the adjective “classical” for notions in probability theory when they are compared with their quantum counterparts, but it should be noted that “classical” does not necessarily mean “old”.)

When the problem of discriminating two probability distributions \( \{p, q\} \) is formulated in Neyman-Pearson’s manner, the asymptotic accuracy of discrimination is represented by the Kullback(-Leibler) divergence (or information)

\[
D(p \parallel q) = E_p \left[ \log \frac{p}{q} \right],
\]

(1)

where \( E_p \) denotes the expectation with respect to the probability distribution \( p \) (Stein’s lemma). On the other hand, for a one-parameter family of probability distributions (1-dimensional statistical model) \( \{p_\theta; \theta \in \Theta(\subset \mathbb{R})\} \), a bound on the estimation accuracy for the parameter \( \theta \) is given in terms of the Fisher information

\[
J(\theta) = E_\theta \left[ \left( \frac{d}{d\theta} \log p_\theta \right)^2 \right] \quad (E_\theta[\cdot] = E_{p_\theta}[\cdot])
\]

(2)

(the Cramér-Rao inequality). As is well known, these quantities are related to each other by

\[
D(p_{\theta + \Delta \theta} \parallel p_\theta) = \frac{1}{2} J(\theta)(\Delta \theta)^2 + o((\Delta \theta)^2). \quad (3)
\]

Generally, we can observe that

parameter estimation for a statistical model \( \{p_\theta; \theta \in \Theta\} \) is almost (at least up to the first order approximation in the limit of \( n \to \infty \)) equivalent to discrimination (hypothesis testing) of \( \{p_{\theta + \Delta \theta}, p_\theta\} \) for an infinitesimal displacement \( \Delta \theta = O(1/\sqrt{n}) \) (where \( n \) is the number of data),

and equation (3) can be regarded as a representation of this observation.

The quantum relative entropy

\[
D(\rho \parallel \sigma) = \text{Tr}[\rho (\log \rho - \log \sigma)],
\]

(4)

which was introduced by Umegaki (1959), is well known as a quantum mechanical analogue of the Kullback divergence. Here \( \rho \) and \( \sigma \) are density operators representing quantum states (see section 3), log is the usual logarithm for operators, and \( \text{Tr} \) denotes the trace. On the other hand, several quantum analogues of Fisher information have been introduced in the study of Cramér-Rao-type inequalities for quantum state estimation ([7, 9, 12, 14]).
Although there are still a lot to be studied on this subject, if we restrict our concern to a one-parameter quantum statistical model \( \{ \rho_\theta ; \theta \in \Theta(\subset \mathbb{R}) \} \), the arguments of the classical estimation theory can be extended to the quantum case in an almost parallel way by adopting the following definition of Fisher information based on the notion of symmetric logarithmic derivative. That is, using the Hermitian operator \( L_\theta \) satisfying
\[
\frac{d}{d\theta} \rho_\theta = \frac{1}{2}(\rho_\theta L_\theta + L_\theta \rho_\theta),
\]
which is called the symmetric logarithmic derivative of the model \( \{ \rho_\theta \} \) at \( \theta \), we define
\[
J(\theta) = \text{Tr}[\rho_\theta (L_\theta)^2].
\]

Now, one may ask whether the same relation as (3) holds between (4) and (6), but in reality we can show that it does not generally hold. What does this mean? One possible answer could be that it means nothing, because almost no result suggesting a relation between the Umegaki entropy \( D \) and a statistical inference problem had been reported so far, and there was no problem in admitting that the \( J(\theta) \) in (6) introduced from an estimation theoretical viewpoint had nothing to do with \( D \). Thus \( D \) and \( J(\theta) \) have been treated separately as notions in different fields, which may seem to be rather strange in comparison with the classical situation around statistics and information theory, so that \( D \) have been studied mostly in connection with operator algebras and quantum statistical mechanics, while \( J(\theta) \) have been positioned in a theory aimed at application to optical communications, etc., and there have been very little interchange between the studies of these two concepts.

Very recently, however, Hiai and Petz [8] showed that the Umegaki entropy is asymptotically obtained as a limit of the Kullback divergence, from which they proved a quantum version of Stein’s lemma (in a slightly weaker form, though). According to this result, the Umegaki entropy \( D(\rho \parallel \sigma) \) represents an asymptotic characteristic of hypothesis testing on two quantum states \( \{ \rho, \sigma \} \), just as the Kullback divergence does. Now that we have this, we cannot simply say any longer that there is no particular meaning in the fact that (3) does not generally hold. In fact, this means that the above-mentioned equivalence between estimation and hypothesis testing in the classical case breaks down in the quantum case. In this paper, we investigate the relation between (4) and (6) by focusing on this breaking down.
2. Kullback Divergence and Fisher Information in the Classical Theory

In this section, we explain some results demonstrating the statistical meanings of the Kullback divergence $D$, the Fisher information $J(\theta)$ and the relation (3) between them in the classical case (i.e., the case of probability distributions), and in addition make some preliminary considerations for later arguments on the quantum case.

We begin with explaining Stein’s lemma. For probability distributions $p, q$ on a finite set $\mathcal{X}$, let us consider the hypothesis testing problem with the null hypothesis $H_0$ and the alternative hypothesis $H_1$ being $p$ and $q$, respectively. A test based on $n$ data is generally represented by a mapping $\varphi(n) : \mathcal{X}^n \rightarrow \{0, 1\}$, and under the i.i.d. (independent and identically distributed) assumption the error probabilities of 1st and 2nd kinds are respectively represented as

$$\alpha(\varphi(n)) = \mathbb{P}_p\{\varphi(n) = 1\} = \mathbb{E}_p[\varphi(n)] = \sum_{(x_1, \ldots, x_n) \in \varphi(n)^{-1}(1)} p(x_1) \cdots p(x_n) \quad (7)$$

and

$$\beta(\varphi(n)) = \mathbb{P}_q\{\varphi(n) = 0\} = 1 - \mathbb{E}_q[\varphi(n)] = \sum_{(x_1, \ldots, x_n) \in \varphi(n)^{-1}(0)} q(x_1) \cdots q(x_n). \quad (8)$$

(Here we do not treat a randomized test.) Letting

$$\beta_*(n; \varepsilon) \overset{\text{def}}{=} \inf \{\beta(\varphi(n)) \mid \varphi(n) : \mathcal{X}^n \rightarrow \{0, 1\}, \alpha(\varphi(n)) \leq \varepsilon\} \quad (9)$$

for an arbitrary $\varepsilon > 0$, Stein’s lemma states that (see, for instance, [5, 6])

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_*(n; \varepsilon) = -D(p \parallel q). \quad (10)$$

On the other hand, in parameter estimation for a statistical model $\{p_\theta : \theta \in \Theta(\subset \mathbb{R})\}$ the Cramér-Rao inequality is important, which states that, when an estimator $T^{(n)} : \mathcal{X}^n \rightarrow \mathbb{R}$ based on $n$ data satisfies the unbiasedness

$$\mathbb{E}_\theta[T^{(n)}] = \theta \quad (\forall \theta \in \Theta),$$

its variance obeys

$$\mathbb{E}_\theta[(T^{(n)} - \theta)^2] \geq 1/nJ(\theta). \quad (11)$$

Even if the unbiasedness is not strictly satisfied, under the consistency condition in the limit of $n \to \infty$ together with some regularity conditions,
we have the asymptotic Cramér-Rao inequality

$$\lim_{n \to \infty} n\mathbb{E}_{\theta}[(T^{(n)} - \theta)^2] \geq 1/J(\theta), \quad (12)$$

for which there exists an estimator (or a sequence of estimators) \( \{T^{(n)}\} \) uniformly attaining the equality for \( \forall \theta \in \Theta \) (e.g. the maximum likelihood estimator).

Now, assuming that the equality holds in (12) and that \( \sqrt{n}(T^{(n)} - \theta) \) is asymptotically distributed according to the normal distribution, the probability density function of \( T^{(n)} \) is approximated by

$$\text{const} \times \exp[-nJ(\theta)(T^{(n)} - \theta)^2/2]$$

when \( n \) is large. On the other hand, it follows from (10) that \( \beta_*(n; \varepsilon) \) is approximated by

$$\text{const} \times \exp[-nD(p \parallel q)].$$

It may be easy to suppose that these facts are related to each other through (3). However, since pursuing this direction in a rigorous manner would require rather complicated arguments, in the sequel we introduce an argument to evaluate the performance of an estimator by the rate of consistency rather than by the mean squared error \( \mathbb{E}_{\theta}[(T^{(n)} - \theta)^2] \), and investigate (3) using this framework.

Let \( \{T^{(n)}\} \) be an arbitrary consistent estimator. It is known (Bahadur [3]; see also Chap. IX of [4]) that for \( \forall \theta \in \Theta \)

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log \mathbb{P}_{\theta}\{|T^{(n)} - \theta| \geq \varepsilon\} \geq -\frac{1}{2}J(\theta) \quad (13)$$

and that when \( \{T^{(n)}\} \) is the maximum likelihood estimator we have

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log \mathbb{P}_{\theta}\{|T^{(n)} - \theta| \geq \varepsilon\} = -\frac{1}{2}J(\theta), \quad (14)$$

where \( \mathbb{P}_{\theta}\{\ldots\} \) denotes the probability of the event \( \{\ldots\} \) with respect to \( p_{\theta} \).

Let us dig a bit deeper into this result in the following.

In order to show (13), it suffices to verify that

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log \mathbb{P}_{\theta}\{T^{(n)} \geq \theta + \varepsilon\} \geq -\frac{1}{2}J(\theta), \quad (15)$$

and for the other direction

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log \mathbb{P}_{\theta}\{T^{(n)} \leq \theta - \varepsilon\} \geq -\frac{1}{2}J(\theta). \quad (16)$$
We actually have, as will be shown below, that for an arbitrarily small \( \varepsilon > 0 \)
\[
\min_{\{T^{(n)}\}} \lim_{n \to \infty} \frac{1}{n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} = -D(p_{\theta+\varepsilon} \| p_\theta) \tag{17}
\]
\[
\min_{\{T^{(n)}\}} \lim_{n \to \infty} \frac{1}{n} \log P_\theta \{ T^{(n)} \leq \theta - \varepsilon \} = -D(p_{\theta-\varepsilon} \| p_\theta), \tag{18}
\]
where the min’s are taken over all consistent estimators. These equations immediately lead to (15) and (16) due to (3).

Equation (17) is proved as follows (and similarly for (18)). First, let us take an arbitrary \( \theta' \in \Theta \) satisfying \( \theta + \varepsilon < \theta' \) and consider a test for the null hypothesis \( H_0 : p_{\theta'} \) and the alternative hypothesis \( H_1 : p_\theta \) as follows; given a data \( (x_1, \ldots, x_n) \), accept \( H_0 \) if \( T^{(n)}(x_1, \ldots, x_n) \geq \theta + \varepsilon \) and accept \( H_1 \) otherwise, where \( \{T^{(n)}\} \) is an arbitrary consistent estimator. Applying Stein’s lemma (10) to this test, we have
\[
\lim_{n \to \infty} \frac{1}{n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} \geq -D(p_{\theta'} \| p_\theta). \tag{19}
\]
Since \( \theta' \) can be arbitrarily close to \( \theta + \varepsilon \), we see that LHS \( \geq \) RHS in (17).

The remaining task for completing the proof of (17) is to show the existence of a consistent estimator \( \{T^{(n)}\} \) satisfying
\[
\lim_{n \to \infty} \frac{1}{n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} = -D(p_{\theta+\varepsilon} \| p_\theta). \tag{20}
\]

Let us first investigate whether (20) is satisfied when \( \{T^{(n)}\} \) is the maximum likelihood estimator (MLE). With this aim, let
\[
L_{\theta'}(x) \overset{\text{def}}{=} \left[ \frac{d}{d\theta} \log p_\theta(x) \right]_{\theta = \theta'} (x \in \mathcal{X})
\]
for each \( \theta' \in \Theta \), and consider the following set consisting of probability distributions on \( \mathcal{X} \):
\[
\Gamma_{\theta'} \overset{\text{def}}{=} \{ q \mid \mathbb{E}_q[L_{\theta'}] = 0 \}.
\]
This set is a mixture family and orthogonally intersects the model at \( p_{\theta'} \) (see [1, 2]). Then, the value of MLE \( T^{(n)}(\vec{x}) \) for a data \( \vec{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n \) is \( \theta' \) if and only if the empirical distribution \( \delta_{\vec{x}} \) determined by \( \vec{x} \) belongs to \( \Gamma_{\theta'} \) [1, 2]. Partitioning the totality of probability distributions on \( \mathcal{X} \) into two parts by the boundary \( \Gamma_{\theta+\varepsilon} \), and letting \( \Delta_{\theta+\varepsilon} \) be the part not including \( p_\theta \) (with \( \Gamma_{\theta+\varepsilon} \) itself included in \( \Delta_{\theta+\varepsilon} \)), it follows that the condition \( T^{(n)}(\vec{x}) \geq \theta + \varepsilon \) on \( \vec{x} \) is (almost) equivalent to \( \delta_{\vec{x}} \in \Delta_{\theta+\varepsilon} \). Therefore, from Sanov’s theorem in the large deviation theory (see e.g. [6, 5]), we have
\[
\lim_{n \to \infty} \frac{1}{n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} = -\min_{q \in \Delta_{\theta+\varepsilon}} D(q \| p_\theta) = -\min_{q \in \Gamma_{\theta+\varepsilon}} D(q \| p_\theta). \tag{21}
\]
If, in addition, the model \( \{p_\theta ; \theta \in \Theta\} \) is an exponential family, the Pythagorean theorem on the divergence\[2] applies to show that for \( \forall q \in \Gamma_{\theta + \varepsilon} \)

\[
D(q \parallel p_\theta) = D(q \parallel p_{\theta + \varepsilon}) + D(p_{\theta + \varepsilon} \parallel p_\theta).
\]

Hence, the RHS of (21) coincides with \(-D(p_{\theta + \varepsilon} \parallel p_\theta)\), and we obtain (20). On the other hand, if the model is not an exponential family, the RHS of (21) > \(-D(p_{\theta + \varepsilon} \parallel p_\theta)\) in general, and therefore the MLE does not satisfy (20). Nevertheless, we can construct a consistent estimator satisfying (20) in the following way. Let us begin with introducing the 1-parameter exponential family \( \{q_\xi ; \xi \in \mathbb{R}\} \) connecting \( p_\theta \) and \( p_{\theta + \varepsilon} \):

\[
q_\xi(x) = A_\xi \{p_\theta(x)\}^{1-\xi} \{p_{\theta + \varepsilon}(x)\}^\xi.
\]

\( A_\xi \triangleq 1/\sum_{x\in\mathcal{X}} \{p_\theta(x)\}^{1-\xi} \{p_{\theta + \varepsilon}(x)\}^\xi \)

(Note that \( q_0 = p_\theta \) and \( q_1 = p_{\theta + \varepsilon} \).) Letting \( S^{(n)} : \mathcal{X}^n \to \mathbb{R} \) denote the MLE of the parameter \( \xi \) of this model and assuming that the true distribution is \( p_{\theta'} \), \( S^{(n)} \) converges to

\[
f(\theta') = \arg\min_\xi D(p_{\theta'} \parallel q_\xi)
\]

in probability as \( n \to \infty \). When \( \varepsilon > 0 \) is sufficiently small, the function \( f \) is one-to-one on an interval including \( [\theta, \theta + \varepsilon] \) and has the inverse function \( g \).

Using this \( g \) we define \( T^{(n)}(\bar{x}) \triangleq g(S^{(n)}(\bar{x})) \), which turns out a consistent estimator for the parameter \( \theta \). In addition, we have

\[
P_\theta\{T^{(n)}(\bar{x}) \geq \theta + \varepsilon\} = P_{q_0}\{S^{(n)} \geq 1\},
\]

and, since \( \{q_\xi ; \xi \in \mathbb{R}\} \) is an exponential family, the previous argument shows that

\[
\lim_{n \to \infty} \frac{1}{n} \log P_{q_0}\{S^{(n)} \geq 1\} = -D(q_1 \parallel q_0) = -D(p_{\theta + \varepsilon} \parallel p_\theta).
\]

It is thus concluded that \( T^{(n)} \) satisfies (20).

From (17) and (18), every consistent estimator \( \{T^{(n)}\} \) satisfies

\[
\lim_{n \to \infty} \frac{1}{n} \log P_\theta\{|T^{(n)} - \theta| \geq \varepsilon\} \geq -\min\{D(p_{\theta + \varepsilon} \parallel p_\theta), D(p_{\theta - \varepsilon} \parallel p_\theta)\},
\]

but there does not generally exist a consistent estimator achieving the equality except for the case when the model is an exponential family. (Note that the minima of (17) and (18) are attained by different \( \{T^{(n)}\}'s \) in general.) If we let \( \{T^{(n)}\} \) be the MLE, it follows from (21) that

\[
\lim_{n \to \infty} \frac{1}{n} \log P_\theta\{|T^{(n)} - \theta| \geq \varepsilon\} = -\min\{C_\theta(\theta + \varepsilon), C_\theta(\theta - \varepsilon)\},
\]

(23)
where
\[ C_\theta(\theta') \stackrel{\text{def}}{=} \min_{q \in \Gamma_{\theta'}} D(q \| p_\theta) \quad (\leq D(p_{\theta'} \| p_\theta)). \] (24)

By a calculation we have
\[ C_\theta(\theta') = -\log E_\theta[e^{tL_{\theta'}}], \] (25)
where \( t \) is the constant determined by
\[ E_\theta[L_{\theta'} e^{tL_{\theta'}}] = 0. \] (26)

It is easy to show that
\[ C_\theta(\theta \pm \varepsilon) = \frac{1}{2} J(\theta) \varepsilon^2 + o(\varepsilon^2), \] (27)
and consequently we see from (23) that (14) holds for the MLE.

3. States and Measurements in Quantum Mechanics

In this section we present a minimum of quantum mechanics needed for later arguments.

Generally, for every quantum mechanical object (a quantum system) there corresponds a Hilbert space (a complex linear space equipped with an inner product), and various physical notions for the system are represented by mathematical notions on \( \mathcal{H} \). The space \( \mathcal{H} \) is generally infinite-dimensional, whereas there are some systems such as spin represented by finite-dimensional spaces. Since rigorous treatment of the infinite-dimensional case would require careful arguments based on the functional analysis, we mainly suppose a finite-dimensional \( \mathcal{H}(\simeq \mathbb{C}^n) \) in the following. Note, however, that most parts can be directly extended to the infinite-dimensional case.

We first explain the physical notion of state and its mathematical representation. A quantum system takes one of various possible states. The notion of state represents comprehensively the statistics of the measurement results for all kinds of measurements which can be performed to the system. In other words, if two quantum systems are in the same state, it means that, when the same measurement is performed to these systems, the measurement results are subject to the same probability distribution. A state is mathematically represented by a linear operator \( \rho : \mathcal{H} \rightarrow \mathcal{H} \) on the Hilbert space \( \mathcal{H} \) satisfying
\[ \rho = \rho^* \geq 0 \quad \text{(Hermitian and positive semidefinite)} \] (28)
and
\[ \text{Tr}[\rho] = 1 \quad (\text{Tr denotes the trace}). \] (29)

Such a \( \rho \) is called a density operator, a statistical operator or a state. The mathematical framework of quantum mechanics consisting of a Hilbert space \( \mathcal{H} \) and a state \( \rho \) on it can be compared with that of probability theory consisting of a measurable space \( (\Omega, \mathcal{F}) \) and a probability measure \( p \) on it.

We next proceed to mathematical representation of quantum measurements. Let us consider an arbitrary measurement, say \( \Pi \), with an arbitrary quantum system (represented by \( \mathcal{H} \)) as the measurement object, and suppose that one of mutually exclusive events \( \{e_1, e_2, \ldots\} \) is observed when the measurement is performed. Generally, we can predict which event is observed only in a probabilistic manner, and the probability distribution is given by
\[ P\{e_i\} = \text{Tr}[\rho \pi_i]. \] (30)

Here \( \rho \) denotes the density operator representing the state of the system (just before performing the measurement), and \( \{\pi_1, \pi_2, \ldots\} = \{\pi_i\} \) is a family of operators on \( \mathcal{H} \) determined by the measurement \( \Pi \) which satisfies the following properties:
\[ \pi_i = \pi_i^* \geq 0 \quad \text{and} \quad \sum_i \pi_i = I \quad (\text{the identity operator}). \] (31)

(It turns out from (28), (29) and (31) that \( P \) in (30) forms a probability distribution, or, in other words, that \( P\{e_i\} \geq 0 \) and \( \sum_i P\{e_i\} = 1 \).) The aspect of the measurement concerning the statistics of the measurement result is thus completely determined by \( \{\pi_i\} \). The other important aspect of a quantum measurement concerning the state change caused by the measurement is not relevant to the subject of the present paper, so we can identify \( \Pi \) with \( \{\pi_i\} \).

In addition, when \( \Pi = \{\pi_i\} \) satisfies
\[ \pi_i \pi_j = \begin{cases} \pi_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \] (32)
\((i.e., \{\pi_i\} \text{ represents the orthogonal projections onto mutually orthogonal subspaces of } \mathcal{H})\) we call \( \Pi \) a simple measurement. This condition means that \( \Pi \) is a pure measurement containing no “additional” uncertainty. Since there is no loss of generality in restricting ourselves to simple measurements in the later arguments, a \( \Pi = \{\pi_i\} \) satisfying (31) and (32) will be simply called a measurement.
In the above definition of measurement, we have not considered the notion of “measurement value”. When the events of measurement results \( \{ e_1, e_2, \ldots \} \) for a measurement \( \Pi = \{ \pi_i \} \) are indicated by distinct real numbers \( \{ x_1, x_2, \ldots \} \), the pair \( \{ \pi_i; x_i \} \) of the measurement \( \{ \pi_i \} \) and the set of measurement values \( \{ x_i \} \) is called an observable. For this \( \{ \pi_i; x_i \} \), a Hermitian operator \( X \) is defined by

\[
X = \sum_i x_i \pi_i. \tag{33}
\]

Conversely, given an arbitrary Hermitian operator \( X \), there uniquely exists an observable \( \{ \pi_i; x_i \} \) satisfying (33), where it turns out that \( \{ x_i \} \) are the eigenvalues of \( X \) and that each \( \pi_i \) is the orthogonal projection \( (\mathcal{H} \to V(x_i)) \) onto the eigenspace \( V(x_i) = \{ \psi \in \mathcal{H} \mid X\psi = x_i\psi \} \). We can thus identify an observable \( \{ \pi_i; x_i \} \) with an Hermitian operator \( X \) by (33). For this observable, letting \( P\{X = x_i\} \) denote the probability that the measurement value \( x_i \) is observed when the state is \( \rho \), we have

\[
P\{X = x_i\} = \text{Tr}[\rho \pi_i], \tag{34}
\]

and the expectation of measurement value is given by

\[
E[X] = \sum_i x_i P\{X = x_i\} = \text{Tr}[\rho X]. \tag{35}
\]

The notion of observables corresponds to that of real-valued random variables in the probability theory.

Finally, we explain about composite systems. When \( n \) quantum systems are represented by Hilbert spaces \( \mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n \), their composite system is represented by the Hilbert space \( \tilde{\mathcal{H}} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n \), where \( \otimes \) denotes the tensor product. Consequently, a state of the composite system is represented by a density operator \( \tilde{\rho} \) on \( \tilde{\mathcal{H}} \). When the element systems constituting the composite system are mutually independent, the state is written in the form \( \tilde{\rho} = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n \).

4. On the Umegaki Relative Entropy

(We assume below for simplicity that density operators \( \rho, \sigma \) are nonsingular in the sense that all the eigenvalues are strictly positive.)

The relative entropy (4) has similar properties to those of the Kullback divergence in many respects[15]. First, it satisfies the following positivity:

\[
D(\rho \parallel \sigma) \begin{cases} > 0 & \text{if } \rho \neq \sigma \\ = 0 & \text{if } \rho = \sigma. \end{cases} \tag{36}
\]
It also satisfies a property corresponding to the monotonicity of the Kullback divergence. Here the monotonicity of the Kullback divergence means that for an arbitrary probability distributions \( p, q \) on a set \( \mathcal{X} \) and for an arbitrary mapping (random variable) \( f : \mathcal{X} \to \mathcal{Y} \) we have \( D(p \parallel q) \geq D(p_f \parallel q_f) \), where \( p_f \) and \( q_f \) respectively denote the probability distributions of \( f \) with respect to \( p \) and \( q \). In order to formulate the general monotonicity of quantum relative entropy it is necessary to introduce the notion of operator algebras and states on them (see [15]), but for the present purpose we need only the special form of monotonicity that for any measurement \( \Pi = \{\pi_i\} \) we have

\[
D(\rho \parallel \sigma) \geq D_\Pi(\rho \parallel \sigma) \overset{\text{def}}{=} D(p \parallel q) \tag{37}
\]

where

\[
p(i) \overset{\text{def}}{=} \text{Tr}[\rho \pi_i], \quad q(i) \overset{\text{def}}{=} \text{Tr}[\sigma \pi_i].
\]

Here \( D_\Pi(\rho \parallel \sigma) \) represents the Kullback divergence for the probability distributions with respect to the measurement \( \Pi \).

In the case when \( \rho \sigma = \sigma \rho \) we have the equality in (37) by using a \( \Pi \) which diagonalizes \( \rho \) and \( \sigma \) simultaneously. In the other case, however, the equality cannot hold: i.e.,

\[
D(\rho \parallel \sigma) > \sup_\Pi D_\Pi(\rho \parallel \sigma) \quad \text{if} \quad \rho \sigma \neq \sigma \rho. \tag{38}
\]

Let us now explain the main result of Hiai and Petz[8] claiming that the inequality in (38) can be replaced with the equality \textit{asymptotically}. For a natural number \( n \), let the tensor product of \( n \) copies of \( \rho \) and \( \sigma \) be denoted by \( \rho^{(n)} \overset{\text{def}}{=} \bigotimes^n \rho \) and \( \sigma^{(n)} \overset{\text{def}}{=} \bigotimes^n \sigma \) respectively. These are density operators on the tensor-product Hilbert space \( \mathcal{H}^{(n)} \overset{\text{def}}{=} \bigotimes^n \mathcal{H} \) (the composite space of \( n \) copies of \( \mathcal{H} \)), and correspond to the classical i.i.d. assumption. It is then easy to show that

\[
D(\rho^{(n)} \parallel \sigma^{(n)}) = n D(\rho \parallel \sigma). \tag{39}
\]

Moreover, it follows from the monotonicity that for an arbitrary measurement \( \Pi^{(n)} \) on \( \mathcal{H}^{(n)} \)

\[
D(\rho^{(n)} \parallel \sigma^{(n)}) \geq D_{\Pi^{(n)}}(\rho^{(n)} \parallel \sigma^{(n)}), \tag{40}
\]

and hence

\[
D(\rho \parallel \sigma) \geq \frac{1}{n} \sup_{\Pi^{(n)}} D_{\Pi^{(n)}}(\rho^{(n)} \parallel \sigma^{(n)}). \tag{41}
\]
Now, letting $n \to \infty$ we have ([8])
\[
D(\rho \parallel \sigma) = \lim_{n \to \infty} \frac{1}{n} \sup_{\Pi(n)} D_{\Pi(n)}(\rho(n) \parallel \sigma(n)).
\] (42)

Remark 1: In the above statement, $\Pi(n)$ is a set of projections corresponding to an arbitrary orthogonal decomposition of the tensor-product space $H^{(n)}$ and includes the special case when $\Pi(n)$ is represented using $n$ measurements $\Pi_1 = \{\pi_{1,i}\}$, $\Pi_2 = \{\pi_{2,i}\}$, ..., $\Pi_n = \{\pi_{n,i}\}$ on $H$ as
\[
\Pi^{(n)} = \left\{ \pi^{(n)}_{i_1i_2\ldots i_n} \right\} \quad \text{where} \quad \pi^{(n)}_{i_1i_2\ldots i_n} = \bigotimes_{k=1}^{n} \pi_{k,i_k},
\] (43)
which is denoted by
\[
\Pi^{(n)} = \bigotimes_{k=1}^{n} \Pi_k.
\] (44)
This type of measurement is physically realized by independently applying measurements $\Pi_1, \ldots, \Pi_n$ one by one to the $n$ systems. A general $\Pi^{(n)}$, on the other hand, is a measurement in which the totality of the $n$ systems is considered as a single system, and includes cases in which quantum mechanical interactions among the element systems are utilized. If we restrict the range of $\sup$ in (41) to those $\Pi^{(n)}$ having the form (44), it turns out that the RHS is $\sup_{\Pi} D_{\Pi}(\rho \parallel \sigma)$, being independent of $n$, and that (42) does not hold.

Furthermore, Hiai and Petz [8] combined (42) with the classical Stein’s lemma to obtain the result mentioned below. Let us consider the hypothesis testing problem about states on the quantum system $H^{(n)} = \bigotimes^n H$ with the two hypotheses $H_0$ : “the true state is $\rho^{(n)} = \bigotimes^n \rho$” and $H_1$ : “the true state is $\sigma^{(n)} = \bigotimes^n \sigma$”. A test in this case is defined as an observable $\varphi^{(n)}$ on $H^{(n)}$ taking values in $\{0, 1\}$. Such a $\varphi^{(n)}$ is a Hermitian operator with its eigenvalues lying in $\{0, 1\}$, which means that it is a projection operator, and $\{\varphi^{(n)}, I - \varphi^{(n)}\}$ forms a binary measurement on $H^{(n)}$. The error probabilities of the 1st and 2nd kinds of the test $\varphi^{(n)}$ are respectively represented as
\[
\alpha(\varphi^{(n)}) = \text{Tr}[\rho^{(n)}\varphi^{(n)}],
\] (45)
\[
\beta(\varphi^{(n)}) = 1 - \text{Tr}[\sigma^{(n)}\varphi^{(n)}].
\] (46)
Letting for an arbitrary $\varepsilon > 0$
\[
\beta_*(n; \varepsilon) \overset{\text{def}}{=} \inf \{\beta(\varphi^{(n)}) \mid \alpha(\varphi^{(n)}) \leq \varepsilon\}
\] (47)
as in (9), we have [8]:
\[
\lim_{n \to \infty} \frac{1}{n} \log \beta_*(n; \varepsilon) \leq -D(\rho \| \sigma),
\]
(48)
\[
\lim_{n \to \infty} \frac{1}{n} \log \beta_*(n; \varepsilon) \geq -\frac{1}{1 - \varepsilon} D(\rho \| \sigma).
\]
(49)
It should be noted that these inequalities are a little weaker in comparison with the equation (10) in the classical case and cannot be regarded as a final result.

5. The Symmetric Logarithmic Derivative and the Quantum Fisher Information

The definition (5) (6) of the Fisher information based on the symmetric logarithmic derivative was introduced by C. W. Helstrom [7] to formulate a version of Cramér-Rao inequality for parameter estimation with respect to a quantum statistical model. In this section we explain some basic results concerning parameter estimation and the Fisher information for a 1-parameter quantum statistical model \( \{\rho_\theta ; \theta \in \Theta (\subset \mathbb{R})\} \), where appropriate regularity conditions on such matters as the smoothness of \( \theta \mapsto \rho_\theta \) are tacitly assumed.

We begin with stating a result corresponding to the classical Cramér-Rao inequality (11) in the case when \( n = 1 \). Generally, an estimator for the parameter \( \theta \) is represented by an observable (i.e. a Hermitian operator) \( T \) on \( \mathcal{H} \). If it satisfies the unbiasedness:
\[
E_\theta[T] \overset{\text{def}}{=} \text{Tr}[\rho_\theta T] = \theta \quad (\forall \theta \in \Theta),
\]
(50)
then we have the following quantum Cramér-Rao inequality:
\[
E_\theta[(T - \theta)^2] = \text{Tr}[\rho_\theta(T - \theta I)^2] \geq 1/J(\theta),
\]
(51)
where \( J(\theta) \) is the (quantum) Fisher information defined by (5) and (6). When we consider \( \rho_\theta^{(n)} = \bigotimes^n \rho_\theta \) instead of \( \rho_\theta \), the corresponding symmetric logarithmic derivative \( L_\theta^{(n)} \) is represented as
\[
L_\theta^{(n)} = L_\theta \otimes I \otimes \cdots \otimes I + I \otimes L_\theta \otimes \cdots \otimes I + \cdots + I \otimes I \otimes \cdots \otimes L_\theta.
\]
It follows that the Fisher information \( J^{(n)}(\theta) \) of the model \( \{\rho_\theta^{(n)}\} \) is
\[
J^{(n)}(\theta) = nJ(\theta),
\]
(52)
and hence we have an inequality similar to (11). (In this case an estimator \( T^{(n)} \) is a Hermitian operator on \( \mathcal{H}^{(n)} \).) In addition, the asymptotic Cramér-Rao inequality similar to (12) also holds.
As with the property (37) of the relative entropy, it holds for the Fisher information that for any measurement \( \Pi = \{ \pi_i \} \)

\[
J(\theta) \geq J_\Pi(\theta) \overset{\text{def}}{=} \mathbb{E}_\theta \left[ \left( \frac{d}{d\theta} \log p_\theta \right)^2 \right], \quad \text{where} \quad p_\theta(i) \overset{\text{def}}{=} \text{Tr}[\rho_\theta \pi_i]. \quad (53)
\]

In contrast to (38), however, we always have [10]

\[
J(\theta) = \max_\Pi J_\Pi(\theta). \quad (54)
\]

Here a \( \Pi \) attaining the max is given by the measurement of the observable represented by the symmetric logarithmic derivative \( L_\theta \) (i.e. the orthogonal projections onto the eigenspaces of \( L_\theta \)). Note that this measurement depends on \( \theta \) in general.

### 6. Relation between the Relative Entropy and the Quantum Fisher Information

In this section we show that a relation similar to (3) does not generally hold in the quantum case, and derive a result to understand the implication of this fact from an estimation-theoretical viewpoint.

Since \( D(\rho_{\theta+\Delta \theta} \parallel \rho_\theta) = O((\Delta \theta)^2) \) holds from (36), we can define a nonnegative quantity \( \tilde{J}(\theta) \) by

\[
D(\rho_{\theta+\Delta \theta} \parallel \rho_\theta) = \frac{1}{2} \tilde{J}(\theta)(\Delta \theta)^2 + o((\Delta \theta)^2). \quad (55)
\]

This is also represented as

\[
\tilde{J}(\theta) = \left[ \frac{\partial^2}{\partial \xi^2} D(\rho_\xi \parallel \rho_\theta) \right]_{\xi=\theta} = \left[ \frac{\partial^2}{\partial \eta^2} D(\rho_\eta \parallel \rho_\theta) \right]_{\eta=\theta} = -\left[ \frac{\partial^2}{\partial \xi \partial \eta} D(\rho_\xi \parallel \rho_\eta) \right]_{\xi=\eta=\theta}.
\]

For any measurement \( \Pi \) on \( \mathcal{H} \) we have \( D(\rho_{\theta+\Delta \theta} \parallel \rho_\theta) \geq D_\Pi(\rho_{\theta+\Delta \theta} \parallel \rho_\theta) \) by the monotonicity, which leads to \( \tilde{J}(\theta) \geq J_\Pi(\theta) \) by (55) and (3). Hence we obtain the following inequality from (54):

\[
\tilde{J}(\theta) \geq J(\theta). \quad (56)
\]

This inequality can be also verified by a direct calculation as follows. First, \( \rho_\theta \) has the spectral decomposition:

\[
\rho_\theta \psi_j(\theta) = \lambda_j(\theta) \psi_j(\theta), \quad (57)
\]

where \( \{ \psi_j(\theta) \} \) are chosen to form an orthogonal normal basis of \( \mathcal{H} \) for every \( \theta \). That is,

\[
\langle \psi_j(\theta) | \psi_k(\theta) \rangle = \delta_{jk}, \quad (58)
\]
where \( \langle \cdot | \cdot \rangle \) denotes the inner product on \( \mathcal{H} \). Using the Dirac notation we have

\[
\rho_\theta = \sum_j \lambda_j(\theta) |\psi_j(\theta)\rangle \langle \psi_j(\theta)|.
\] (59)

It then follows from (28) (29) that the eigenvalues \( \{\lambda_j(\theta)\} \) are nonnegative real numbers and satisfy \( \sum_j \lambda_j(\theta) = 1 \). Note that there may be some equal values in \( \{\lambda_j(\theta)\} \). We further assume that \( \lambda_j(\theta) \) and \( \psi_j(\theta) \) are smooth functions of \( \theta \). In this situation, with the aid of

\[
\log \rho_\theta = \sum_j \log \lambda_j(\theta) |\psi_j(\theta)\rangle \langle \psi_j(\theta)|
\]

and

\[
L_\theta = \sum_j \sum_k \frac{2 \langle \psi_j(\theta)| (d\rho_\theta/d\theta) \psi_k(\theta) \rangle}{\lambda_j(\theta) + \lambda_k(\theta)} |\psi_j(\theta)\rangle \langle \psi_k(\theta)|,
\]

we can show the following relations:

\[
\tilde{J}(\theta) = \sum_j \lambda_j \left\{ \frac{d}{d\theta} \log \lambda_j \right\}^2 + 2 \sum_j \sum_k \lambda_j \left( \log \frac{\lambda_j}{\lambda_k} \right) a_{jk}
\] (60)

\[
J(\theta) = \sum_j \lambda_j \left\{ \frac{d}{d\theta} \log \lambda_j \right\}^2 + 4 \sum_j \sum_k \lambda_j \left( \frac{\lambda_j - \lambda_k}{\lambda_j + \lambda_k} \right)^2 a_{jk},
\] (61)

where

\[
a_{jk} \overset{\text{def}}{=} \left| \left\langle \frac{d\psi_j}{d\theta} | \psi_k \right\rangle \right|^2.
\] (62)

Noting that \( a_{jk} = a_{kj} \) follows from (58), inequality (56) can be derived from (60) and (61). We also see that the equality does not generally hold in (56).

The results given below are considered to correspond to (15)-(18) in the classical theory, which enable us to understand the meaning of inequality (56) from an estimation-theoretical viewpoint. From here on we treat sequences \( \{T^{(n)}\} \) of observables \( T^{(n)} \) on \( \mathcal{H}^{(n)} \) which are regarded as (sequences of) estimators. Then the probability distribution of \( T^{(n)} \) with respect to the state \( \rho_\theta^{(n)} = \bigotimes^n \rho_\theta \) is determined. When it holds for this distribution that

\[
\lim_{n \to \infty} \mathbb{P}_\theta \{ |T^{(n)} - \theta| \geq \varepsilon \} = 0, \quad (\forall \theta \in \Theta, \forall \varepsilon > 0)
\] (63)

we call \( \{T^{(n)}\} \) a consistent estimator.
**Theorem 2:** For \( \forall \theta \in \Theta \) and \( \forall \varepsilon > 0 \),

\[
\inf_{\{T^{(n)}\}} \lim_{n \to \infty} \frac{1}{n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} = -D(\rho_{\theta+\varepsilon} \mid \rho_\theta),
\]

(64)

\[
\inf_{\{T^{(n)}\}} \lim_{n \to \infty} \frac{1}{n} \log P_\theta \{ T^{(n)} \leq \theta - \varepsilon \} = -D(\rho_{\theta-\varepsilon} \mid \rho_\theta),
\]

(65)

where the inf’s are taken over all consistent estimators.

From Theorem 2 and (55) we obtain the following.

**Corollary 3:** For \( \forall \theta \in \Theta \),

\[
\lim_{\varepsilon \to 0} \inf_{\{T^{(n)}\}} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} = -\frac{1}{2} \tilde{J}(\theta),
\]

(66)

\[
\lim_{\varepsilon \to 0} \inf_{\{T^{(n)}\}} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{ T^{(n)} \leq \theta - \varepsilon \} = -\frac{1}{2} \tilde{J}(\theta).
\]

(67)

**Theorem 4:** For \( \forall \theta \in \Theta \),

\[
\inf_{\{T^{(n)}\}} \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} = -\frac{1}{2} J(\theta),
\]

(68)

\[
\inf_{\{T^{(n)}\}} \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{ T^{(n)} \leq \theta - \varepsilon \} = -\frac{1}{2} J(\theta),
\]

(69)

where the inf’s are tentatively considered to be taken over all consistent estimators \( \{T^{(n)}\} \), whereas there remains a certain problem as will be mentioned in the remark after the proof.

**Proof of Theorem 2:** Here we prove (64) only, with ‘LHS \( \geq \) RHS’ shown first. Let \( \{T^{(n)}\} \) be an arbitrary consistent estimator, and let

\[
p_n \stackrel{\text{def}}{=} P_\theta \{ T^{(n)} \geq \theta + \varepsilon \}, \quad q_n \stackrel{\text{def}}{=} P_\theta' \{ T^{(n)} \geq \theta + \varepsilon \}
\]

for arbitrary \( \varepsilon > 0 \) and \( \theta' > \theta + \varepsilon \). From the consistency of \( \{T^{(n)}\} \) we have

\[
\lim_{n \to \infty} p_n = 0, \quad \lim_{n \to \infty} q_n = 1.
\]

(70)

In addition, it follows from the monotonicity of the Kullback divergence that

\[
D_{T^{(n)}}(\rho^{(n)}_{\theta'} \mid \rho^{(n)}_{\theta}) \geq q_n \log \frac{q_n}{p_n} + (1 - q_n) \log \frac{1 - q_n}{1 - p_n}
\]

\[
= -q_n \log p_n - (1 - q_n) \log(1 - p_n) - h(q_n),
\]

(71)
where \( D_{T^{(n)}}(\rho_{\theta'}^{(n)} \parallel \rho_{\theta}^{(n)}) \) denotes the Kullback divergence for the probability distributions of \( T^{(n)} \) with respect to the states \( \rho_{\theta'}^{(n)} \) and \( \rho_{\theta}^{(n)} \), and \( h(q_n) \) \( \overset{\text{def}}{=} \) \(-q_n \log q_n - (1 - q_n) \log(1 - q_n)\). Noting that (71) leads to

\[
\log p_n \geq \frac{1}{q_n} \left\{ D_{T^{(n)}}(\rho_{\theta'}^{(n)} \parallel \rho_{\theta}^{(n)}) + (1 - q_n) \log(1 - q_n) + h(q_n) \right\},
\]

we obtain from (70)

\[
\lim_{n \to \infty} \frac{1}{n} \log p_n \geq -\lim_{n \to \infty} \frac{1}{n} D_{T^{(n)}}(\rho_{\theta'}^{(n)} \parallel \rho_{\theta}^{(n)}).
\]

(72)

Now, the monotonicity of the relative entropy yields

\[
D_{T^{(n)}}(\rho_{\theta'}^{(n)} \parallel \rho_{\theta}^{(n)}) \leq D(\rho_{\theta'}^{(n)} \parallel \rho_{\theta}^{(n)}) = nD(\rho_{\theta'} \parallel \rho_{\theta}),
\]

and therefore from the above inequality we have

\[
\lim_{n \to \infty} \frac{1}{n} \log p_n \geq -D(\rho_{\theta'} \parallel \rho_{\theta}).
\]

Since \( \theta' \) is arbitrary within \( \theta' > \theta + \varepsilon \), we obtain

\[
\lim_{n \to \infty} \frac{1}{n} \log p_n \geq -D(\rho_{\theta+\varepsilon} \parallel \rho_{\theta}).
\]

‘LHS \( \geq \) RHS’ in (64) has thus been proved.

We next proceed to show ‘LHS \( \leq \) RHS’. Fix \( \theta \) and \( \varepsilon > 0 \) arbitrarily. Then for any \( \delta > 0 \) it follows from (42) that there exist a natural number \( l \) and a measurement \( \Pi^{(l)} = \{\pi_i^{(l)}\} \) on \( \mathcal{H}^{(l)} \) satisfying

\[
\frac{1}{l} D_{\Pi^{(l)}}(\rho_{\theta+\varepsilon}^{(l)} \parallel \rho_{\theta}^{(l)}) = \frac{1}{l} D(\rho_{\theta+\varepsilon}^{(l)} \parallel \rho_{\theta}^{(l)}) \geq D(\rho_{\theta+\varepsilon} \parallel \rho_{\theta}) - \delta,
\]

(73)

where \( \rho_{\theta}^{(l)} \) denotes the probability distribution defined by \( \rho_{\theta}^{(l)}(i) = \text{Tr}[\rho_{\theta}(\pi_i^{(l)})] \). Using this measurement, let us construct for each \( n \geq l \) an observable \( T^{(n)} \) on \( \mathcal{H}^{(n)} \) in the following way. First, let \( k = k_n \) and \( j = j_n \) be natural numbers determined by \( n = lk + j \), \( 0 \leq j \leq l - 1 \). Regarding \( \mathcal{H}^{(n)} \) as the composite system of \( k \) copies of \( \mathcal{H}^{(l)} \) and one \( \mathcal{H}^{(j)} \), the operation to perform the measurement \( \Pi^{(l)} \) to each \( \mathcal{H}^{(l)} \) and no measurement to \( \mathcal{H}^{(j)} \) defines a measurement \( \Pi^{(n)} \) on \( \mathcal{H}^{(n)} \). That is, \( \Pi^{(n)} = \{\pi_i^{(n)}\} \) is defined by

\[
\pi_i^{(n)} = \pi_i^{(l)} \otimes \cdots \otimes \pi_i^{(l)} \otimes I^{(j)}.
\]

(74)

Concerning the statistical model \( \{p_{\theta}^{(l)}\} \) consisting of the probability distributions \( p_{\theta}^{(l)} \), let \( \ell^{(k)} = \ell^{(k)}(i_1, \ldots, i_k) \) be an estimator for \( \theta \), based on \( k \) data
Proof of Theorem 4: We prove (68) only, beginning with showing ‘LHS ≥ RHS’. As was shown in (72), for any consistent estimator \( \{T^{(n)}\} \) and \( \forall \varepsilon > 0, \forall \theta' > \theta + \varepsilon \), we have

\[
\frac{1}{n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} \geq - \lim_{n \to \infty} \frac{1}{n} D_{T^{(n)}}(\rho^{(n)}_{\theta'}) \| \rho^{(n)}_\theta),
\]

which yields

\[
\frac{1}{n} \log P_\theta \{ T^{(n)} \geq \theta + \varepsilon \} \geq - \lim_{n \to \infty} \frac{1}{n} D_{T^{(n)}}(\rho^{(n)}_{\theta+\varepsilon}) \| \rho^{(n)}_\theta) \]

\[
= - \lim_{n \to \infty} \frac{1}{n} J_{T^{(n)}}(\theta).
\]

(See the remark below for \((*)\).) Here \( J_{T^{(n)}}(\theta) \) denotes the Fisher information of the statistical model consisting of the probability distributions of \( T^{(n)} \) with respect to the states \( \rho^{(n)}_\theta \), and the last equality follows from (3). Since by (52) and (53)

\[
J_{T^{(n)}}(\theta) \leq J^{(n)}(\theta) = nJ(\theta),
\]

‘LHS ≥ RHS’ in (68) is proved by (78).

We next show ‘LHS ≤ RHS’. For an arbitrarily fixed \( \theta \), let \( \Pi = \{\pi_i\} \) be a measurement on \( \mathcal{H} \) which attains the max in (54); i.e.,

\[
J(\theta) = J_\Pi(\theta).
\]
Let \( t^{(n)}(i_1, \ldots, i_n) \) be a consistent estimator (e.g. the maximum likelihood estimator) for the model \( \{p_\theta\} \) consisting of the probability distributions \( p_\theta(i) \overset{\text{def}}{=} \text{Tr}[\rho_\theta \pi_i] \) which attains the lower bound of (15) (\( -\frac{1}{2} J(\theta) \) in the present case), and define

\[
T^{(n)} \overset{\text{def}}{=} \sum_{i_1 \cdots i_n} t^{(n)}(i_1, \ldots, i_n) \pi_{i_1} \otimes \cdots \otimes \pi_{i_n}.
\]  

(80)

Then \( T^{(n)} \) is an observable on \( \mathcal{H}^{(n)} \) and satisfies

\[
\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{ T^{(n)} \geq \theta + \epsilon \} = -\frac{1}{2} J(\theta) = -\frac{1}{2} J_\Pi(\theta).
\]  

(81)

‘LHS \leq \text{RHS}' has thus been proved.

**Remark 5:** In the proof above we have changed the order of \( \lim_{n \to \infty} \) and \( \lim_{\epsilon \downarrow 0} \) at (*) in (78), but its validity is not necessarily clear. Actually we need to investigate whether the validity follows from the consistency of \( T^{(n)} \) or otherwise what kind of assumption is necessary for the validity. The main point is, however, whether inequality (78) itself, i.e.

\[
\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{ T^{(n)} \geq \theta + \epsilon \} \geq -\frac{1}{2} \lim_{n \to \infty} \frac{1}{n} J^{(n)}_T(\theta),
\]

is valid or not. So let us tentatively consider the set \( \mathcal{C} \) of consistent estimators satisfying this inequality and replace \( \inf \{ T^{(n)} \} \) with \( \inf \{ T^{(n)} \} \in \mathcal{C} \) in (64)-(69) all at once. Then it is clear that ‘LHS \geq \text{RHS}' holds in these equations, but in fact ‘LHS \leq \text{RHS}' also holds. That is, the inf’s in (64)-(69) are attained in the range \( \{ T^{(n)} \} \in \mathcal{C} \). The reason is as follows. In the classical case, we can show that for a consistent estimator satisfying a certain regularity condition we always have

\[
\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\epsilon^2 n} \log P_\theta \{ T^{(n)} \geq \theta + \epsilon \} = -\frac{1}{2} \lim_{n \to \infty} \frac{1}{n} J^{(n)}_T(\theta).
\]  

(82)

The estimators (75) and (80) constructed to show ‘LHS \leq \text{RHS}' in the proofs of (64) and (68) are both directly derived from classical estimators and satisfy (82). Therefore these estimators belong to \( \mathcal{C} \).

7. Discussions

Comparing (66) (67) with (68) (69), the difference between them lies in the positions of \( \lim_{\epsilon \downarrow 0} \) and \( \inf \{ T^{(n)} \} \), which immediately implies inequality (56). The results may be given an intuitive explanation as follows; equations (66) and (68) are both evaluating error probabilities in discriminating \( \theta \) and \( \theta + \epsilon \).
for infinitesimally small \( \varepsilon > 0 \), but they differ in that in (66) optimization (of \( \{T^{(n)}\} \)) is made for each \( \varepsilon \), while in (68) a kind of ‘uniform’ optimization for all infinitesimal \( \varepsilon > 0 \) is considered, so the condition of the latter is more restrictive than that of the former to yield a larger error probability. In the classical case, we can also consider both kinds of optimizations, but it turns out that they coincide with each other owing to (3). In the quantum case, on the other hand, they differ in general. The difference takes an extreme form in the case when the elements \( \rho_{\theta} \) of the model are pure states. We say that \( \rho_{\theta} \) is a pure state when its rank is 1, which is characterized by the condition \( \rho_{\theta}^2 = \rho_{\theta} \) (under (28) and (29)). In this case there is only one 1 in the eigenvalues \( \{\lambda_j(\theta)\} \) of \( \rho_{\theta} \) with the rest being 0. Hence it follows from (60) and (61) that \( J(\theta) < \infty = \tilde{J}(\theta) \).

There is also a remarkable difference between the substance of the inf’s in (66) (67) and that in (68) (69). As mentioned in Remark 1 in section 4, if we only consider measurements on \( \mathcal{H}^{(n)} \) of the form given in (44), then (42) breaks down and so do (64)-(67). On the other hand, in the proof of Th. 2 the estimator \( T^{(n)} \) in (80) is constructed from the measurement \( \Pi^{(n)} = \bigotimes^n \Pi \), where \( \Pi \) is a measurement on \( \mathcal{H} \) satisfying (79), and this \( T^{(n)} \) attains the inf in (68). Therefore, the result remains the same in (68) and (69) even if we restrict ourselves to measurements on \( \mathcal{H}^{(n)} \) of the form (44).

Corresponding to (13) and (14), it is obvious from (66)-(69) that

\[
\lim_{\varepsilon \downarrow 0} \inf_{\{T^{(n)}\}} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{|T^{(n)} - \theta| \geq \varepsilon\} \geq -\frac{1}{2} \tilde{J}(\theta),
\]

(83)

\[
\inf_{\{T^{(n)}\}} \lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{|T^{(n)} - \theta| \geq \varepsilon\} \geq -\frac{1}{2} J(\theta).
\]

(84)

The equality in (83) does not generally hold, while as for (84), letting \( t^{(n)} \) in constructing \( T^{(n)} \) by (80) be the maximum likelihood estimator, we have

\[
\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{|T^{(n)} - \theta| \geq \varepsilon\} = -\frac{1}{2} J(\theta),
\]

and therefore

\[
\inf_{\{T^{(n)}\}} \lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} \log P_\theta \{|T^{(n)} - \theta| \geq \varepsilon\} = -\frac{1}{2} J(\theta).
\]

(85)

Although the above-mentioned results are all evaluations at a fixed \( \theta \), from a proper estimation-theoretical standpoint it is more desirable to have uniform evaluations for \( \theta \). In the classical case, the maximum likelihood estimator attains the equalities in (13), (15) and (16) uniformly (i.e. for all
\( \theta \). In the quantum case, it is also expected that an estimator uniformly attaining the inf’s in (68), (69) and (85) can be constructed by modifying the construction (80) with application of the method [11,13] in which the maximum likelihood estimation is carried out together with an adaptive search of a measurement approximating the \( \Pi \) in (79). It is impossible, however, to apply the same argument to (66) and (67).

References


CHAPTER 28

Two Quantum Analogues of Fisher Information
from a Large Deviation Viewpoint
of Quantum Estimation

Masahito Hayashi

Abstract. We discuss two quantum analogues of Fisher information, symmetric logarithmic derivative (SLD) Fisher information and Kubo-Mori-Bogoljubov (KMB) Fisher information from a large deviation viewpoint of quantum estimation and prove that the former gives the true bound and the latter gives the bound of consistent superefficient estimators. In another comparison, it is shown that the difference between them is characterized by the change of the order of limits.

1. Introduction

Fisher information not only plays a central role in statistical inference, but also coincides with a natural inner product in a distribution family. It is defined as

\[ J_\theta := \int_\Omega l_\theta(\omega)^2 p_\theta(\omega) \, d\omega, \quad l_\theta(\omega) p_\theta(\omega) = \frac{dp_\theta(\omega)}{d\theta} \]  

for a probability distribution family \( \{p_\theta|\theta \in \Theta \subset \mathbb{R}\} \) with a probability space \( \Omega \). However, the quantum version of Fisher information cannot be uniquely determined. In general, there is a serious arbitrariness concerning the order among non-commutative observables in the quantization of products of several variables. The problem of the arbitrariness of the quantum version of Fisher information is due to the same reason. The geometrical properties of its quantum analogues have been discussed by many authors [1, 34, 35, 36].

One quantum analogue is the Kubo-Mori-Bogoljubov (KMB) Fisher information \( \tilde{J}_\rho \) defined by

\[ \tilde{J}_\theta := \int_0^1 \text{Tr} \rho_\theta^t \tilde{L}_\theta \rho_\theta^{1-t} \tilde{L}_\theta \, dt, \quad \int_0^1 \rho_\theta^t \tilde{L}_\theta \rho_\theta^{1-t} \, dt = \frac{d\rho_\theta}{d\theta} \]  

This chapter is reprinted from *J. Phys. A: Math. and Gen.*, 35, 7689-7727, 2002. Since large part of Appendices consist of the technical proofs, they are omitted. They are available at quant-ph/0202003.
for a quantum state family \( \{ \rho_\theta \in S(H) | \theta \in \Theta \} \), where \( S(H) \) is the set of density matrices on \( H \) and the Hilbert space \( H \) corresponds to the physical system of interest [1, 34, 35, 36]. As proven in B, it can be characterized as the limit of quantum relative entropy, which plays an important role in several topics of quantum information theory, for example, quantum channel coding [24, 38], quantum source coding [16, 37, 42] and quantum hypothesis testing [22, 33]. Moreover, as mentioned in section 3, this inner product is closely related to the canonical correlation of the linear response theory in statistical mechanics [27]. As mentioned in A, it appears to be the most natural quantum extension from an information geometrical viewpoint. Thus, one might expect that it is significant in quantum estimation, but its estimation-theoretical characterization has not been sufficiently clarified.

Another quantum analogue is symmetric logarithmic derivative (SLD) Fisher information

\[
J_\theta := \text{Tr} L_\theta^2 \rho_\theta, \quad \frac{1}{2} (L_\theta \rho_\theta + \rho_\theta L_\theta) = \frac{d \rho_\theta}{d \theta},
\]

(3)

where \( L_\theta \) is called the symmetric logarithmic derivative [20]. It is closely related to the achievable lower bound of mean square error (MSE) not only for the one-parameter case [20, 21, 23], but also for the multi-parameter case [11, 13, 14] in quantum estimation. The difference between the two can be regarded as the difference in the order of the operators, and reflects the two ways of defining Fisher information for a probability distribution family.

Currently, the former is closely related to the quantum information theory while the latter is related to the quantum estimation theory. These two inner products have been discussed only in separate contexts. In this paper, to clarify the difference between them, we introduce a large deviation viewpoint of quantum estimation as a unified viewpoint, whose classical version was initiated by Bahadur [2, 3, 4]. This method may not be conventional in mathematical statistics, but seems a suitable setting for a comparison between two quantum analogues from an estimation viewpoint. This type of comparison was initiated by Nagaoka [31, 32], and is discussed in further depth in this paper. Such a large deviation evaluation of quantum estimation is closely related to the exponent of the overflow probability of quantum universal variable-length coding [19].

This paper is structured as follows: Before we state the main results, we review the classical estimation theory including Bahadur’s large deviation theory, which has been done in section 2. After this review, we briefly outline the main results in section 3, i.e., the difference is characterized from three
contexts. To simplify the notations, even if we need the Gauss notation \([\ ]\), we omit it when this does not cause confusion. Some proofs are very complicated and are presented in the appendixes.

2. Review of Classical Estimation Theory

We review the relationship between the parameter estimation for the probability distribution family \(\{p_\theta | \theta \in \Theta \subset \mathbb{R}\}\) with a probability space \(\Omega\) and its Fisher information. The definition of Fisher information is given not only by (1), but also by the limit of the relative entropy (Kullback-Leibler divergence) \(D(p \parallel q) := \int_\Omega (\log p(\omega) - \log q(\omega)) p(\omega) \, d\omega\) as

\[
J_\theta := \lim_{\varepsilon \to 0} \frac{2}{\varepsilon^2} D(p_{\theta + \varepsilon} \parallel p_{\theta}).
\]

These two definitions (1) and (4) coincide under some regularity conditions for a family.

Next, we consider a map \(f\) from \(\Omega\) to \(\Omega'\). Similarly to other information quantities, (for example Kullback divergence etc) the inequality \(J_\theta \geq J'_\theta\) (5) holds, where \(J'_\theta\) is Fisher information of the family \(\{p_\theta \circ f^{-1} | \theta \in \Theta\}\). Inequality (5) is called the monotonicity. According to Čencov [8], any information quantities satisfying (5) coincide with a constant times Fisher information \(J_\theta\).

For an estimator that is defined as a map from the data set \(\Omega\) to the parameter set \(\Theta\), we sometimes consider the unbiasedness condition:

\[
\int_\Omega T(\omega)p_\theta(\omega) \, d\omega = \theta, \quad \forall \theta \in \Theta.
\]

The MSE of any unbiased estimator \(T\) is evaluated by the following inequality (Cramér-Rao inequality),

\[
\int_\Omega (T(\omega) - \theta)^2 p_\theta(\omega) \, d\omega \geq \frac{1}{J_\theta},
\]

which follows from Schwartz inequality with respect to (w.r.t.) the inner product \(\langle X, Y \rangle := \int_\Omega X(\omega)Y(\omega)p_\theta(\omega) \, d\omega\) for variables \(X, Y\). When the number of data \(\vec{\omega}_n := (\omega_1, \ldots, \omega_n)\), which obeys the unknown probability \(p_\theta\), is sufficiently large, we discuss a sequence \(\{T_n\}\) of estimators \(T_n(\vec{\omega}_n)\). If \(\{T_n\}\) is suitable as a sequence of estimators, we can expect that it converges.
to the true parameter $\theta$ in probability, i.e., it satisfies the weak consistency condition:

$$\lim_{n \to \infty} p^n_\theta \{ |T_n - \theta| > \varepsilon \} = 0, \quad \forall \varepsilon > 0, \forall \theta \in \Theta.$$  \hfill (8)

Usually, the performance of a sequence $\{T_n\}$ of estimators is measured by the speed of its convergence. As one criterion, we focus on the speed of the convergence in MSE. If a sequence $\{T_n\}$ of estimators satisfies the weak consistency condition and some regularity conditions, the asymptotic version of Cramér-Rao inequality,

$$\lim_{n \to \infty} n \int_\Omega (T_n(\bar{\omega}_n) - \theta)^2 p^n_\theta(\omega) \, d\omega \geq \frac{1}{J_\theta},$$  \hfill (9)

holds. If it satisfies only the weak consistency condition, it is possible that it surpasses the bound of (9) at a specific subset. Such a sequence of estimators is called superefficient. We can reduce its error to any amount at a specific subset with the measure 0 under the weak consistency condition (8).

As another criterion, we evaluate the decreasing rate of the tail probability:

$$\beta(\{T_n\}, \theta, \varepsilon) := \lim_{n \to \infty} \frac{-1}{n} \log p^n_\theta \{ |T_n - \theta| > \varepsilon \}. \hfill (10)$$

This method was initiated by Bahadur [2, 3, 4], and was a much discussed topic among mathematical statisticians in the 1970’s. From the monotonicity of the divergence, we can prove the inequality

$$\beta(\{T_n\}, \theta, \varepsilon) \leq \min \{ D(p_{\theta+\varepsilon} || p_\theta), D(p_{\theta-\varepsilon} || p_\theta) \}$$  \hfill (11)

for any weakly consistent sequence $\{T_n\}$ of estimators. Its proof is essentially given in our proof of Theorem 5. Since it is difficult to analyze $\beta(\{T_n\}, \theta, \varepsilon)$ except in the case of an exponential family, we focus on another quantity $\alpha(\{T_n\}, \theta) := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\{T_n\}, \theta, \varepsilon)$. For an exponential family, see Appendix K. Taking the limit $\varepsilon \to +0$, we obtain the inequality

$$\alpha(\{T_n\}, \theta) \leq \frac{J_\theta}{2}.$$  \hfill (12)

If $T_n$ is the maximum likelihood estimator (MLE), the equality of (12) holds under some regularity conditions for the family [4, 9]. This type of discussion is different from the MSE type of discussion in deriving (12) from only the weak consistency condition. Therefore, there is no consistent superefficient estimator w.r.t. the large deviation evaluation.

Indeed, we can relate the above large deviation type of discussion in the estimation to Stein’s lemma in simple hypothesis testing as follows. In
simple hypothesis testing, we decide whether the null hypothesis should be accepted or rejected from the data \( \tilde{\omega}_n := (\omega_1, \ldots, \omega_n) \) which obeys an unknown probability. For the decision, we must define an acceptance region \( A_n \) as a subset of \( \Omega^n \). If the null hypothesis is \( p \) and the alternative is \( q \), the first error (though the true distribution is \( p \), we reject the null hypothesis) probability \( \beta_{1,n}(A_n) \) and the second error (though the true distribution is \( q \), we accept the null hypothesis) probability \( \beta_{2,n}(A_n) \) are given by

\[
\beta_{1,n}(A_n) := 1 - p^n(A_n), \quad \beta_{2,n}(A_n) := q^n(A_n).
\]

Regarding the decreasing rate of the second error probability under the constant constraint of the first error probability, the equation

\[
\lim_{n \to \infty} \frac{-1}{n} \log \min \{ \beta_{2,n}(A_n) | \beta_{1,n}(A_n) \leq \varepsilon \} = D(p\|q), \quad \varepsilon > 0
\]

holds (Stein’s lemma). Inequality (11) can be derived from this lemma. We can regard the large deviation type of evaluation in the estimation to be Stein’s lemma in the case where the null hypothesis is close to the alternative one.

3. Outline of Main Results

Let us return to the quantum case. In a quantum setting, we focus two quantum analogues of Fisher information, KMB Fisher information \( \tilde{J}_\theta \) and SLD Fisher information \( J_\theta \). Indeed, if the state \( \rho_\theta \) is degenerate, SLD \( L_\theta \) is not uniquely determined. However, as is proven in Appendix C, SLD Fisher information \( J_\theta \) is uniquely determined, i.e., it is independent of the choice of the SLD \( L_\theta \).

On the other hand, according to Chap. 7 in Amari and Nagaoka [1], \( \tilde{L}_\theta \) has another form

\[
\tilde{L}_\theta = \frac{d \log \rho_\theta}{d\theta}.
\]  

As is proven by using formula (14) in Appendix B, KMB Fisher information \( \tilde{J}_\theta \) can be characterized as the limit of the the quantum relative entropy \( D(\rho\|\sigma) := \text{Tr} \rho (\log \rho - \log \sigma) \) in the following way

\[
\tilde{J}_\theta = \lim_{\varepsilon \to 0} \frac{2}{\varepsilon^2} D(\rho_{\theta + \varepsilon}\|\rho_\theta).
\]

Moreover, in the linear response theory of statistical physics, given an equilibrium state \( \rho \), when a variable \( A \) fluctuates with a small value \( \delta \), another
variable $B$ also is thought to fluctuate with a constant times $\delta$ [27]. Its coefficient is called the canonical correlation and given by

$$
\int_0^1 \text{Tr} \rho_\theta^t (A - \text{Tr} \rho A) \rho_\theta^{1-t} (B - \text{Tr} \rho B) \, dt.
$$

Thus, KMB Fisher information $\tilde{J}_\theta$ is thought to be more natural from a viewpoint of statistical physics.

As another quantum analogue, right logarithmic derivative (RLD) Fisher information $\tilde{J}_\theta$:

$$
\tilde{J}_\theta := \text{Tr} \rho_\theta \tilde{L}_\theta \tilde{L}_\theta^*, \quad \rho_\theta \tilde{L}_\theta = \frac{d\rho_\theta}{d\theta}
$$

is known. When $\rho_\theta$ does not commute $\frac{d\rho_\theta}{d\theta}$ and $\rho_\theta > 0$, the RLD $\tilde{L}_\theta$ is not self-adjoint. Since it is not useful in the one-parameter case, we do not discuss it in this paper. Since the difference in definitions can be regarded as the difference in the order of operators, these quantum analogues coincide when all states of the family are commutative with each other. However, in the general case, they do not coincide and the inequality $\tilde{J}_\theta \geq J_\theta$ holds, as exemplified in section 4. Concerning some information-geometrical properties, see Appendix A.

In the following, we consider how the roles these quantum analogues of Fisher information play in the parameter estimation for the state family. As is discussed in detail in section 4, the estimator is described by the pair of positive operator valued measure (POVM) $M$ (which corresponds to the measurement and is defined in section 4) and the map from the data set to the parameter space $\Theta$. Similarly to the classical case, we can define an unbiased estimator. For any unbiased estimator $E$, the SLD Cramér-Rao inequality

$$
V(E) \geq \frac{1}{J_\theta}
$$

holds, where $V(E)$ is the mean square error (MSE) of the estimator $E$.

In an asymptotic setting, as a quantum analogue of the $n$-i.i.d. condition, we treat the quantum $n$-i.i.d. condition, i.e., we consider the case where the number of systems independently prepared in the same unknown state is sufficiently large, in section 5. In this case, the measurement is denoted by a POVM $M^n$ on the composite system $\mathcal{H}^\otimes n$ and the state is described by the tensor product density matrix $\rho^\otimes n$. Of course, such POVMs include a POVM that requires quantum correlations between the respective quantum systems in the measurement apparatus. Similarly to the classical case,
for a sequence $\vec{E} = \{E^n\}$ of estimators, we can define the weak consistency condition given in (31). In mathematical statistics, the square root $n$ consistency, local asymptotic minimax theorems and Bayesian theorem are important topics as the asymptotic theory, but it seems too difficult to link these quantum settings and KMB Fisher information $\tilde{J}_\theta$. Thus, in this paper, in order to compare two quantum analogues from a unified framework, we adopt Bahadur’s large deviation theory as follows. As is discussed in section 5, we can similarly define the quantities $\beta(\vec{E}, \theta, \varepsilon), \alpha(\vec{E}, \theta)$. Similarly to (11)(12), under the weak consistency (WC) condition, the inequalities

\[ \beta(\vec{E}, \theta, \varepsilon) \leq \min \{ D(\rho_{\theta+\varepsilon} || \rho_{\theta}), D(\rho_{\theta-\varepsilon} || \rho_{\theta}) \} \]
\[ \alpha(\vec{E}, \theta) \leq \frac{1}{2} \tilde{J}_\theta \quad (18) \]

hold. From these discussions, the bound in the large deviation type of evaluation seems different from the one in the MSE case. However, as mentioned in section 6, the inequality

\[ \alpha(\vec{E}, \theta) \leq \frac{1}{2} J_\theta \quad (19) \]

holds if the sequence $\vec{E}$ satisfies the strong consistency (SC) condition introduced in section 6 as a stronger condition. As is mentioned in section 7, these bounds can be attained in their respective senses. Therefore, roughly speaking, the difference between the two quantum analogues can be regarded as the difference in consistency conditions and can be characterized as

\[ \sup_{E\text{SC}} \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\vec{E}, \theta, \varepsilon) = \frac{1}{2} J_\theta, \quad \sup_{E\text{WC}} \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\vec{E}, \theta, \varepsilon) = \frac{1}{2} \tilde{J}_\theta. \]

Even if we restrict our estimators to strongly consistent ones, the difference between two appears as

\[ \sup_{\vec{M}\text{SC}} \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\vec{M}, \theta, \varepsilon) = \frac{J_\theta}{2} \quad (20) \]
\[ \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \sup_{\vec{M}\text{SC}} \beta(\vec{M}, \theta, \varepsilon) = \frac{\tilde{J}_\theta}{2}, \quad (21) \]

where, for a precise statement, as expressed in section 9, we need more complicated definitions.

However, we should consider that the bound $\frac{J_\theta}{2}$ is more meaningful for the following two reasons. The first reason is the fact that we can construct the sequence of estimators attaining the bound $\frac{J_\theta}{2}$ at all points, which is
proven in section 7. On the other hand, there is a sequence of estimators attaining the bound $\frac{j_\theta}{2}$ at one point $\theta$, but it cannot attain the bound at all points. The other reason is the naturalness of the conditions for deriving the bound $\frac{j_\theta}{2}$. In other words, an estimator attaining $\frac{j_\theta}{2}$ is natural, but an estimator attaining $\frac{\tilde{j}_\theta}{2}$ is very irregular. Such a sequence of estimators can be regarded as a consistent super-efficient estimator and does not satisfy regularity conditions other than the weak consistency condition. This type of discussion of the super-efficiency is different from the MSE type of discussion in that any consistent super-efficient estimator is bounded by inequality (18).

To consider the difference between the two quantum analogues of Fisher information in more details, we must analyze how we can achieve the bound $\frac{\tilde{j}_\theta}{2}$. It is important in this analysis to consider the relationship between the above discussion and the quantum version of Stein’s lemma in simple hypothesis testing. Similarly to the classical case, when the null hypothesis is the state $\rho$ and the alternative is the state $\sigma$, we evaluate the decreasing rate of the second error probability under the constant constraint $\varepsilon > 0$ of the first error probability. As was proven in quantum Stein’s lemma, its exponential component is given by the quantum relative entropy $D(\rho\|\sigma)$ for any $\varepsilon > 0$. Hiai and Petz [22] constructed a sequence of tests to attain the optimal rate $D(\rho\|\sigma)$, by constructing the sequence $\{M^n\}$ of POVMs such that

$$\lim_{n\to\infty} \frac{1}{n}D(P^{M^n}_\rho \| P^{M^n}_\sigma) = D(\rho\|\sigma).$$

(22)

Ogawa and Nagaoka [33] proved that there is no test exceeding the bound $D(\rho\|\sigma)$. It was proven by Hayashi [15] that by using the group representation theory, we can construct the POVM satisfying (22) independently of $\rho$. For the reader’s convenience, we give a review of this in Appendix J. As discussed in section 7.2, this type of construction is useful for the construction of an estimator attaining the bound $\frac{j_\theta}{2}$ at one point. Since the proper bound of the large deviation is $\frac{j_\theta}{2}$, we cannot regard the quantum estimation as the limit of the quantum Stein’s lemma.

In order to consider the properties of estimators attaining the bound $\frac{j_\theta}{2}$ at one point from another viewpoint, we consider the restriction that makes such a construction impossible. We introduce a class of estimators whose POVMs do not require a quantum correlation in the quantum apparatus in section 8. In this class, we assume that the POVM on the $l$-th system is chosen from $l-1$ data. We call such an estimator an adaptive estimator. When an adaptive estimator $\tilde{E}$ satisfies the weak consistency condition, the
inequality
\[ \alpha(\vec{E}, \theta) \leq \frac{1}{2} J_\theta \] (23)
holds (See section 6). Similarly, we can define a class of estimators that use quantum correlations up to \( m \) systems. We call such an estimator an \( m \)-adaptive estimator. For any \( m \)-adaptive weakly consistent estimator \( \vec{E} \), inequality (23) holds. Therefore, it is impossible to construct a sequence of estimators attaining the bound \( \frac{J_\theta}{2} \) if we fix the number of systems in which we use quantum correlations. As mentioned in section 8, taking limit \( m \to \infty \), we obtain
\[
\lim_{m \to \infty} \lim_{\varepsilon \to 0} \sup_{M : m \text{-AWC}} \frac{1}{\varepsilon^2} \beta(M, \theta, \varepsilon) = \frac{J_\theta}{2},
\] (24)
where \( m \)-AWC denotes an \( m \)-adaptive weakly consistent estimator. However, as the third characterization of the difference between the two quantum analogues, as precisely mentioned in section 9, the equation
\[
\lim_{\varepsilon \to 0} \lim_{m \to \infty} \sup_{M : m \text{-ASC}} \frac{1}{\varepsilon^2} \beta(M, \theta, \varepsilon) = \frac{\tilde{J}_\theta}{2}
\] (25)
holds, where \( m \)-ASC denotes an \( m \)-adaptive strongly consistent estimator. A more narrow class of estimators is treated in equation (25) than in equation (21). Equations (24) and (25) indicate that the order of limits \( \lim_{m \to \infty} \) and \( \lim_{\varepsilon \to 0} \) is more crucial than the difference between two types of consistencies.

Remark 1: In the estimation only of the spectrum of a density matrix in a unitary-invariant family, the natural inner product in the parameter space is unique and equals Fisher inner product in the distribution family whose element is the probability distribution corresponding to eigenvalues of a density matrix. In addition, the achievable bound is derived by Keyl and Werner [26], and coincides with the bound uniquely given by the above inner product. For detail, see Appendix L.

4. Review of Non-Asymptotic Setting in Quantum Estimation

In a quantum system, in order to discuss the probability distribution which the data obeys, we must define a POVM.

A POVM \( M \) is defined as a map from Borel sets of the data set \( \Omega \) to the set of bounded, self-adjoint and positive semi-definite operators, which
satisfies
\[ M(\phi) = 0, \quad M(\Omega) = 1, \quad \sum_i M(B_i) = M(\cup B_i) \] for disjoint sets.

If the state on the quantum system \( \mathcal{H} \) is a density operator \( \rho \) and we perform a measurement corresponding to a POVM \( M \) on the system, the data obeys the probability distribution \( P^M_\rho(B) := \text{Tr} \, \rho M(B) \). If a POVM \( M \) satisfies \( (M(B))^2 = M(B) \) for any Borel set \( B \), \( M \) is called a projection-valued measure (PVM). The spectral measure of a self-adjoint operator \( X \) is a PVM, and is denoted by \( E(X) \). For \( 1 > \lambda > 0 \) and any POVMs \( M_1 \) and \( M_2 \) taking values in \( \Omega \), the POVM \( B \mapsto \lambda M_1(B) + (1 - \lambda)M_2(B) \) is called the random combination of \( M_1 \) and \( M_2 \) in the ratio \( \lambda : 1 - \lambda \). Even if \( M_1 \)'s data set \( \Omega_1 \) is different from \( M_2 \)'s data set \( \Omega_2 \), \( M_1 \) and \( M_2 \) can be regarded as POVMs taking values in the disjoint union set \( \Omega_1 \coprod \Omega_2 := (\Omega_1 \times \{1\}) \cup (\Omega_2 \times \{2\}) \). In this case, we can define a random combination of \( M_1 \) and \( M_2 \) as a POVM taking values in \( \Omega_1 \coprod \Omega_2 \) and call it the disjoint random combination.

In this paper, we simplify the probability \( P^M_\rho(\theta) \) and the relative entropies \( D(\rho_\theta \| \rho_{\theta_0}) \) and \( D(P^M_\rho(\theta_0 \| \theta_1)) \) to \( P^M_\theta \), \( D(\theta_0 \| \theta_1) \) and \( D^M(\theta_0 \| \theta_1) \), respectively.

In the one-parameter quantum estimation, the estimator is described by a pair comprising a POVM and a map from its data set to the real number set \( \mathbb{R} \). Since the POVM \( M \circ T^{-1} \) takes values in the real number set \( \mathbb{R} \), we can regard any estimator as a POVM taking values in the real number set \( \mathbb{R} \). In order to evaluate MSE, Helstrom [20, 21] derived the SLD Cramér-Rao inequality as a quantum counterpart of Cramér-Rao inequality (29). If an estimator \( M \) satisfies
\[ \int_{\mathbb{R}} x \text{Tr} \, \rho \theta M(\,dx\,) = \theta, \quad \forall \theta \in \Theta, \quad (26) \]
it is called unbiased. If \( \theta - \theta_0 \) is sufficiently small, we can obtain the following approximation in the neighborhood of \( \theta_0 \):
\[ \int_{\mathbb{R}} x \text{Tr} \, \rho_{\theta_0} M(\,dx\,) + \left( \int_{\mathbb{R}} x \text{Tr} \frac{\partial \rho_\theta}{\partial \theta} \bigg|_{\theta = \theta_0} M(\,dx\,) \right)(\theta - \theta_0) \cong \theta_0 + (\theta - \theta_0). \]

It implies the following two conditions:
\[ \int_{\mathbb{R}} x \text{Tr} \frac{\partial \rho_\theta}{\partial \theta} \bigg|_{\theta = \theta_0} M(\,dx\,) = 1 \quad (27) \]
\[ \int_{\mathbb{R}} x \text{Tr} \rho_{\theta_0} M(\,dx\,) = \theta_0. \quad (28) \]
If an estimator $M$ satisfies (27) and (28), it is called locally unbiased at $\theta_0$.

For any locally unbiased estimator $M$ (at $\theta$), the inequality, which is called the SLD Cramér-Rao inequality,

$$\int_{\mathbb{R}} (x - \theta)^2 \text{Tr} \rho_\theta M(\,d x) \geq \frac{1}{J_\theta}$$

holds. Similarly to the classical case, this inequality is derived from the Schwartz inequality with respect to SLD Fisher information $\langle X|Y \rangle := \text{Tr} \rho_\theta \frac{X^Y + X_Y}{2}$ [20, 21, 23].

The equality of (29) holds when the estimator is given by the spectral decomposition $E\left(\frac{L_\theta}{J_\theta} + \theta\right)$ of $\frac{L_\theta}{J_\theta} + \theta$, where $L_\theta$ is the SLD at $\theta$ and is defined by (3). This implies that SLD Fisher information $J_{\theta_0}$ coincides with Fisher information at $\theta_0$ of the probability family $\left\{P_\theta^{E\left(\frac{L_\theta}{J_\theta} + \theta_0\right)} \mid \theta \in \Theta\right\}$.

The monotonicity of quantum relative entropy [29, 39] gives the following evaluation of the probability family $\left\{P_\theta^{E\left(\frac{L_\theta}{J_\theta} + \theta_0\right)} \mid \theta \in \Theta\right\}$:

$$D_{E\left(\frac{L_\theta}{J_\theta} + \theta_0\right)}(\theta||\theta_0) \leq D(\theta||\theta_0).$$

Taking the limit $\theta \to \theta_0$, we have

$$J_\theta \leq \tilde{J}_\theta.$$

In this paper, we discuss inequality (30) from the viewpoint of the large deviation type of evaluation of the quantum estimation. The following families are treated as simple examples of the one-parameter quantum state family, in the latter.

**Example 2: [One-parameter equatorial spin 1/2 system state family]:**

$$S_r := \left\{\rho_\theta := \frac{1}{2} \begin{pmatrix} 1 + r \cos \theta & r \sin \theta \\ r \sin \theta & 1 - r \cos \theta \end{pmatrix} \mid 0 \leq \theta < 2\pi\right\}$$

In this family, we calculate

$$D(\rho_\theta||\rho_0) = \frac{r}{2} (1 - \cos \theta) \log \frac{1 + r}{1 - r}, \quad \tilde{J}_\theta = \frac{r}{2} \log \frac{1 + r}{1 - r}, \quad J_\theta = r^2.$$
Example 3: [One-parameter quantum Gaussian state family and half-line quantum Gaussian state family]: We define the boson coherent vector $|\alpha\rangle := e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$, where $|n\rangle$ is the number vector on $L^2(\mathbb{R})$. The quantum Gaussian state is defined as

$$\rho_\theta := \frac{1}{\pi N} \int_{\mathbb{C}} |\alpha\rangle \langle e^{-|\alpha-\theta|^2/2} d^2\alpha, \quad \forall \theta \in \mathbb{C}. $$

We call $\{\rho_\theta|\theta \in \mathbb{R}\}$ the one-parameter quantum Gaussian state family, and call $\{\rho_\theta|\theta \geq 0|\theta \in \mathbb{R}^+ = [0, \infty)\}$ the half-line quantum Gaussian state family. In this family, we can calculate

$$D(\rho_\theta||\rho_{\theta_0}) = \log \left( 1 + \frac{1}{N} \right) |\theta - \theta_0|^2, \quad J_\theta = 2 \log \left( 1 + \frac{1}{N} \right), \quad \tilde{J}_\theta = \frac{2}{N + \frac{1}{2}}. $$

5. The Bound under the Weak Consistency Condition

We introduce the quantum independent-identical density (i.i.d.) condition in order to treat an asymptotic setting. Suppose that $n$-independent physical systems are prepared in the same state $\rho$. Then, the quantum state of the composite system is described by

$$\rho^\otimes n := \rho \otimes \cdots \otimes \rho \text{ on } \mathcal{H}^\otimes n, $$

where the tensor product space $\mathcal{H}^\otimes n$ is defined by

$$\mathcal{H}^\otimes n := \mathcal{H} \otimes \cdots \otimes \mathcal{H}. \quad \text{We call this condition the quantum i.i.d. condition, which is a quantum analogue of the independent-identical distribution condition. In this setting, any estimator is described by a POVM $M^n$ on $\mathcal{H}^\otimes n$, whose data set is } \mathbb{R}. \text{ In this paper, we simplify } P_{\rho_\theta^\otimes n}^{M^n} \text{ and } D(P_{\rho_\theta^\otimes n}^{M^n}||P_{\rho_{\theta_0}^\otimes n}^{M^n}) \text{ to } P_{\rho_\theta}^{M^{n}} \text{ and } D^{M^{n}}(\theta_0||\theta_1). \text{ The notation } M \times n \text{ denotes the POVM in which we perform the POVM } M \text{ for the respective } n \text{ systems.}

\textbf{Definition 4: [Weak consistency condition]: A sequence of estimators } \bar{M} := \{M^n\}_{n=1}^{\infty} \text{ is called weakly consistent if}

$$\lim_{n \to \infty} P_{\rho_\theta}^{M^n} \left\{ |\hat{\theta} - \theta| > \varepsilon \right\} = 0, \quad \forall \theta \in \Theta, \forall \varepsilon > 0, \quad (31)$$

where $\hat{\theta}$ is the estimated value.
This definition means that the estimated value $\hat{\theta}$ converges to the true value $\theta$ in probability, and can be regarded as the quantum extension of (8).

Now, we focus on the exponential component of the tail probability as follows:

$$\beta(\vec{M}, \theta, \varepsilon) := \limsup_{n \to \infty} -\frac{1}{n} \log P_{\theta}^M \left\{ |\hat{\theta} - \theta| > \varepsilon \right\}.$$  

We usually discuss the following value instead of $\beta(\vec{M}, \theta, \varepsilon)$

$$\alpha(\vec{M}, \theta) := \limsup_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\vec{M}, \theta, \varepsilon)$$  

because it is too difficult to discuss $\beta(\vec{M}, \theta, \varepsilon)$. The following theorem can be proven from the monotonicity of the quantum relative entropy.

**Theorem 5:** (Nagaoka [31,32]) If a POVM $M^n$ on $\mathcal{H}^{\otimes n}$ satisfies the weakly consistent condition (31), the inequalities

$$\beta(\vec{M}, \theta, \varepsilon) \leq \inf_{\varepsilon' \in \Theta} D(\rho_{\theta'} \| \rho_{\theta}) \quad \text{for } |\theta' - \theta| > \varepsilon, \quad \alpha(\vec{M}, \theta) \leq \frac{\tilde{J}_\theta}{2}$$  

hold.

Even if the parameter set $\Theta$ is not open (e.g., the closed half-line $\mathbb{R}^+ := [0, \infty)$), this theorem holds.

**Proof:** The monotonicity of the quantum relative entropy yields the inequality

$$D(\rho_{\theta'}^{\otimes n} \| \rho_{\theta}^{\otimes n}) \geq p_{n,\theta'} \log \frac{p_{n,\theta'}}{p_{n,\theta}} + (1 - p_{n,\theta'}) \log \frac{1 - p_{n,\theta'}}{1 - p_{n,\theta}},$$  

for any $\theta'$ such that satisfies $|\theta' - \theta| > \varepsilon$, where we denote the probability $P_{\theta'}^M \left\{ |\hat{\theta} - \theta| > \varepsilon \right\}$ by $p_{n,\theta'}$. Using the inequality

$$-(1 - p_{n,\theta'}) \log (1 - p_{n,\theta}) \geq 0,$$

we have

$$-\frac{\log P_{\theta}^M \left\{ |\hat{\theta} - \theta| > \varepsilon \right\}}{n} = -\frac{\log p_{n,\theta}}{n} \leq \frac{D(\rho_{\theta'}^{\otimes n} \| \rho_{\theta}^{\otimes n}) + h(p_{n,\theta'})}{np_{n,\theta'}},$$  

where $h$ is the binary entropy defined by $h(x) := -x \log x - (1 - x) \log (1 - x)$. Since the assumption guarantees that $p_{n,\theta'} \to 1$, the inequality

$$\beta(\vec{M}, \theta, \varepsilon) \leq D(\rho_{\theta'} \| \rho_{\theta})$$  

holds, where we use the additivity of quantum relative entropy:

$$D(\rho_{\theta'}^{\otimes n} \| \rho_{\theta}^{\otimes n}) = nD(\rho_{\theta'} \| \rho_{\theta}).$$
Thus, we obtain (33). Taking the limit $\varepsilon \to 0$ in inequality (36), we obtain (34).

As another proof, we can prove this inequality as a corollary of the quantum Stein’s lemma [22, 33].

6. The Bound under the Strong Consistency Condition

As discussed in section 4, the SLD Cramér-Rao inequality guarantees that the lower bound of MSE is given by SLD Fisher information. Therefore, it is expected that the bound is connected with SLD Fisher information for large deviation. In order to discuss the relationship between SLD Fisher information and the bound for large deviation, we need another characterization with respect to the limit of the tail probability. We thus define

$$\beta(\bar{M}, \theta, \varepsilon) := \liminf_{n \to \infty} -\frac{1}{n} \log p^{M^n}_\theta \left\{ |\hat{\theta} - \theta| > \varepsilon \right\}$$

$$\alpha(\bar{M}, \theta) := \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\bar{M}, \theta, \varepsilon).$$

(37)

In the following, we attempt to link the quantity $\alpha(\bar{M}, \theta)$ with SLD Fisher information. For this purpose, it is suitable to focus on an information quantity that satisfies the additivity and the monotonicity, as in the proof of Theorem 5. Its limit should be SLD Fisher information. The Bures distance $b(\rho, \sigma) := \sqrt{2(1 - \text{Tr} |\sqrt{\rho} \sqrt{\sigma}|)} = \sqrt{\min_{U: \text{unitary}} \text{Tr}(\sqrt{\rho} - \sqrt{\sigma}U)(\sqrt{\rho} - \sqrt{\sigma}U)^*}$ is known to be an information quantity whose limit is SLD Fisher information, as mentioned in Lemma 6. Of course, it can be regarded as a quantum analogue of the Hellinger distance, and satisfies the monotonicity.

Lemma 6: (Uhlmann [40], Matsumoto [30]). If there exists an SLD $L_\theta$ satisfying (3), then the equation

$$\frac{1}{4} J_\theta = \lim_{\varepsilon \to 0} \frac{b^2(\rho_\theta, \rho_{\theta+\varepsilon})}{\varepsilon^2}$$

(38)

holds.

A proof of Lemma 6 is given in Appendix C. As discussed in the latter, the Bures distance satisfies the monotonicity. Unfortunately, the Bures distance does not satisfy the additivity.

However, the quantum affinity $I(\rho \parallel \sigma) := -8 \log \text{Tr} |\sqrt{\rho} \sqrt{\sigma}| = -8 \log (1 - \frac{1}{2} b(\rho, \sigma)^2)$ satisfies the additivity:

$$I(\rho^\otimes n \parallel \sigma^\otimes n) = n I(\rho \parallel \sigma).$$

(39)
Its classical version is called affinity in the following form [28]:

\[ I(p\|q) = -8 \log \left( \sum_i \sqrt{p_i} \sqrt{q_i} \right) \]  \tag{40}

As a trivial deformation of (38), the equation

\[ \lim_{\varepsilon \to 0} \frac{I(\rho_\theta \| \rho_\theta + \varepsilon)}{\varepsilon^2} = J_\theta \]  \tag{41}

holds. The quantum affinity satisfies the monotonicity w.r.t. any measurement \( M \) (Jozsa [25], Fuchs [10]):

\[ I(\rho \| \sigma) \geq I(P^M_\rho \| P^M_\sigma) = -8 \log \sum_\omega \left( \sqrt{P^M_\rho(\omega)} \sqrt{P^M_\sigma(\omega)} \right) \]  \tag{42}

The most simple proof of (42) is given by Fuchs [10] who directly proved that

\[ \text{Tr} \sqrt{\rho \sigma} \sqrt{\rho} \leq \sum_\omega \left( \sqrt{P^M_\rho(\omega)} \sqrt{P^M_\sigma(\omega)} \right) \]  \tag{43}

For the reader’s convenience, a proof of (43) is given in D. From (39),(41) and (42), we can expect that SLD Fisher information is, in a sense, closely related to a large deviation type of bound. From the additivity and the monotonicity of the quantum affinity, we can show the following lemma.

**Lemma 7:** The inequality

\[ 4 \inf_{\{s|1 \geq s \geq 0\}} \left( \beta'(\widetilde{M}, \theta, s\delta) + \beta'(\widetilde{M}, \theta + \delta, (1 - s)\delta) \right) \leq I(\rho_\theta \| \rho_{\theta + \delta}) \]  \tag{44}

holds, where we define \( \beta'(\widetilde{M}, \theta, \delta) := \lim_{\varepsilon \to +0} \beta(\widetilde{M}, \theta - \varepsilon) \).

A proof of Lemma 7 is given in Appendix E. However, Lemma 7 cannot yield an inequality w.r.t. \( \alpha(\widetilde{M}, \theta) \) under the weak consistency condition, unlike inequality (36). Therefore, we consider a stronger condition, which is given in the following.

**Definition 8:** [Strong consistency condition]: A sequence of estimators \( \widetilde{M} = \{M^n\}_{n=1}^\infty \) is called strongly consistent if the convergence of (37) is uniform for the parameter \( \theta \) and if \( \alpha(\widetilde{M}, \theta) \) is continuous for \( \theta \). A sequence of estimators is called strongly consistent at \( \theta \) if there exists a neighborhood \( U \) of \( \theta \) such that it is strongly consistent in \( U \).
The square root \( n \) consistency is familiar in the field of mathematical statistics. However, in the large deviation setting, this strong consistency seems more suitable than the square root \( n \) consistency.

As a corollary of Lemma 7, we have the following theorem.

**Theorem 9:** Assume that there exists the SLD \( L_\theta \) satisfying (3). If a sequence of estimators \( \vec{M} = \{M^n\}_{n=1}^\infty \) is strongly consistent at \( \theta \), then the inequality

\[
\alpha(\vec{M}, \theta) \leq \frac{J_\theta}{2}
\]  

holds.

**Proof:** From the above assumption, for any real \( \varepsilon > 0 \) and any element \( \theta \in \Theta \), there exists a sufficiently small real \( \delta > 0 \) such that \( (\alpha(\vec{M}, \theta) - \varepsilon)\varepsilon' \leq \beta'(\vec{M}, \theta, \varepsilon') \beta'(\vec{M}, \theta + \delta, \varepsilon') \) for \( \forall \varepsilon' < \delta \). Therefore, inequality (44) yields the relations

\[
2(\alpha(\vec{M}, \theta) - \varepsilon)\delta^2 = 4(\alpha(\vec{M}, \theta) - \varepsilon) \inf_{\{s|1 \geq s \geq 0\}} \left( s^2 \delta^2 + (1 - s)^2 \delta^2 \right)
\]

\[
\leq 4 \inf_{\{s|1 \geq s \geq 0\}} \left( \beta'(\vec{M}, \theta, s\delta) + \beta'(\vec{M}, \theta + \delta, (1 - s)\delta) \right) \leq I(\rho_\theta || \rho_{\theta+\delta}). \tag{46}
\]

Lemmas 6 and (46) guarantee (45) for \( \forall \theta \in \Theta \).

**Remark 10:** Inequality (43) can be regarded as a special case of the monotonicity w.r.t. any trace-preserving CP (completely positive) map \( C : S(\mathcal{H}_1) \to S(\mathcal{H}_2) \):

\[
(\text{Tr} |\sqrt{\rho} \sqrt{\sigma}|)^2 \leq \left( \text{Tr} \left| \sqrt{C(\rho)} \sqrt{C(\sigma)} \right| \right)^2
\]

which is proven by Jozsa [25] because the map \( \rho \mapsto P_\rho^M \) can be regarded as a trace-preserving CP map from the \( C^* \) algebra of bounded operators on \( \mathcal{H} \) to the commutative \( C^* \) algebra \( C(\Omega) \), where \( \Omega \) is the data set.

### 7. Attainabilities of the Bounds

Next, we discuss the attainabilities of the two bounds \( \tilde{J}_\theta \) and \( J_\theta \) in their respective senses. In this section, we discuss the attainabilities in two cases: the first case is the one-parameter quantum Gaussian state family, and the second case is an arbitrary one-parameter finite-dimensional quantum state family that satisfies some assumptions.
7.1. One-parameter quantum Gaussian state family

In this subsection, we discuss the attainabilities in the one-parameter quantum Gaussian state family.

**Theorem 11:** In the one-parameter quantum Gaussian state family, the sequence of estimators $\vec{M}_s = \{M^{s,n}\}_{n=1}^\infty$ (defined in the following) satisfies the strong consistency condition and the relations

$$\alpha(\vec{M}^s, \theta) = \frac{J_\theta}{2} = \frac{1}{N + \frac{1}{2}}.$$ (48)

[Construction of $\vec{M}^s$]: We perform the POVM $E(Q)$ for all systems, where $Q$ is the position operator on $L^2(\mathbb{R})$. The estimated value $\xi_n$ is determined to be the mean value of $n$ data. ■

**Proof:** Since the equation

$$P_{E(Q)}(\alpha)\langle\alpha| dx) = \sqrt{\frac{2}{\pi}} e^{-2(x-\alpha)^2} dx$$

holds, we have the equation

$$P_{\theta}^{E(Q)}(dx) = \left( P_{\rho_0}^{E(Q)}(dx) \right) = \frac{1}{\pi N} \int_{\mathbb{C}} P_{|\alpha\rangle\langle\alpha|}^{E(Q)}(dx) e^{-\frac{|\alpha - \theta|^2}{N}} d^2\alpha$$

$$= \sqrt{\frac{2}{\pi(2N+1)}} e^{-\frac{2(x-\theta)^2}{2N+1}} dx.$$ 

Thus, we obtain the equation

$$P_{\theta}^{M^{s,n}}(d\xi_n) = \sqrt{\frac{2}{\pi(2N+1)n}} e^{-\frac{2(\xi_n-\theta)^2}{(2N+1)n}} d\xi_n,$$

which implies that

$$\beta(\vec{M}^s, \theta, \varepsilon) = \lim_{n} -\frac{1}{n} \log P_{\theta}^{M^{s,n}}\{|\xi_n - \theta| > \varepsilon\} = \frac{\varepsilon^2}{N + \frac{1}{2}}.$$ (49)

Therefore, the sequence of estimators $\vec{M}^s = \{M^{s,n}\}_{n=1}^\infty$ attains the bound $J_\theta/2$ and satisfies the strong consistency condition. ■

**Proposition 12:** In the half-line quantum Gaussian state family, the sequence of estimators $\vec{M}^w = \{M^{w,n}\}_{n=0}^\infty$ (defined in the following) satisfies the weak consistency condition and the strong consistency condition at
The relations
\[
\alpha(M^w, 0) = \alpha(M^w, 0) = \frac{J_0}{2} = \log \left(1 + \frac{1}{N}\right),
\]
\[
\alpha(M^w, \theta) = \alpha(M^w, \theta) = \frac{J_\theta}{2} = \frac{1}{N + \frac{1}{2}}, \quad \forall \theta \in \mathbb{R}^+ \setminus \{0\}.
\]

This proposition indicates the significance of the uniformity of the convergence of (37). This proposition is proven in Appendix G.

[Construction of \(M^w\):] We perform the following unitary evolution:
\[
\rho_\theta^{\otimes n} \mapsto \rho_{\sqrt{n}\theta} \otimes \rho_0^{\otimes (n-1)}.
\]

For detail, see Appendix F. We perform the number measurement \(E(N)\) of the first system whose state is \(\rho_{\sqrt{n}\theta}\), and let \(k\) be its data, where the number operator \(N\) is defined as \(N := \sum_n n|n\rangle\langle n|\). The estimated value \(T_n\) is determined by \(T_n := \sqrt{\frac{k}{n}}\).

**Theorem 13:** In the one-parameter quantum Gaussian state family, for any \(\theta \in \mathbb{R}\), the sequence of estimators \(M^w_{\theta_1} = \{M^w_{\theta_1,n}\}_{n=1}^\infty\) (defined in the following) satisfies the weak consistency condition and the relations
\[
\alpha(M^w_{\theta_1}, \theta_1) = \alpha(M^w_{\theta_1}, \theta_1) = \frac{J_\theta}{2} = \log \left(1 + \frac{1}{N}\right).
\]

[Construction of \(M^w_{\theta_1}\):] We divide \(n\) systems into two groups. One consists of \(\sqrt{n}\) systems and the other, of \(n - \sqrt{n}\) systems. We perform the PVM \(E(Q)\) for every system in the first group. Let \(\xi_{\sqrt{n}}\) be the mean value in the first group, i.e., we perform the PVM \(M_{s, \sqrt{n}}\) for the first system. At the second step, we perform the following unitary evolution for the second group.
\[
\rho_\theta^{\otimes (n-\sqrt{n})} \mapsto \rho_{\theta - \theta_1}^{\otimes (n-\sqrt{n})}
\]

For details, see Appendix F. We perform the POVM \(M^{w,n-\sqrt{n}}\) for the system whose state is \(\rho_{\theta - \theta_1}^{\otimes (n-\sqrt{n})}\); the data is written as \(T_{n-\sqrt{n}}\). Then, we decide the final estimated value \(\hat{\theta}\) as
\[
\hat{\theta} := \theta_1 + \text{sgn}(\xi_{\sqrt{n}} - \theta_1)T_{n-\sqrt{n}}.
\]

**Proof:** Since
\[
P_{\theta_1}^{M_{\theta_1,n}} \left\{ |\hat{\theta} - \theta_1| > \varepsilon \right\} = P_{0}^{M^{w,n-\sqrt{n}}} \left\{ |T_{n-\sqrt{n}}| > \varepsilon \right\},
\]
we have
\[
\beta(M_{\theta_1}^w, \theta_1) = \lim_{n \to \infty} -\frac{1}{n} \log P_{\theta_1}^{M_{\theta_1}^w, n} \left\{ \left| \hat{\theta} - \theta_1 \right| > \varepsilon \right\} 
\]
\[
= \lim_{n \to \infty} -\frac{n - \sqrt{n}}{n} \log P_0^{M_{\theta_1}^w, \sqrt{n}} \left\{ \left| T_{n-\sqrt{n}} \right| > \varepsilon \right\} = \beta(M_{\theta_1}^w, 0).
\]
As is shown in Appendix G, we have
\[
\beta(M_{\theta_1}^w, 0) = \varepsilon^2 \log \left( 1 + \frac{1}{N} \right),
\]
which implies (52). Next, we prove the consistency in the case where \( \theta > \theta_1 \).
In this case, it is sufficient to discuss the case where \( \theta - \theta_1 > \varepsilon > 0 \). Since the first measurement \( M_{\theta_1}^\sqrt{n} \) and the second one \( M_{\theta_1}^w, \sqrt{n} \) are performed independently, we obtain
\[
P_{\theta_1}^{M_{\theta_1}^w, n} \left\{ \left| \hat{\theta} - \theta_1 \right| > \varepsilon \right\} \leq P_{\theta}^{M_{\theta_1}^w, \sqrt{n}} \left\{ \left| T_{n-\sqrt{n}} - (\theta - \theta_1) \right| > \varepsilon \right\}
\]
\[
+ P_{\theta}^{M_{\theta_1}^w, \sqrt{n}} \left\{ \xi_\sqrt{n} - \theta_1 \leq 0 \right\}.
\]
Proposition 12 guarantees that the first term goes to 0, and Theorem 11 guarantees that the second term goes to 0. Thus, we obtain the consistency of \( M_{\theta_1}^w \). Similarly, we can prove the weak consistency the case where \( \theta < \theta_1 \).

7.2. Finite dimensional family

In this subsection, we treat the case where the dimension of the Hilbert space \( \mathcal{H} \) is \( k \) (finite). As for the attainability of the RHS of inequality (45), we have the following lemma.

**Lemma 14:** Let \( \theta_0 \) be fixed in \( \Theta \). Under Assumptions 1 and 2, the sequence of estimators \( \tilde{M}_{\theta_0}^s \) (defined in the following) satisfies the strong consistency condition at \( \theta_0 \) (defined in Def. 8) and the relation
\[
\alpha(\tilde{M}_{\theta_0}^s, \theta_0) = \alpha(\tilde{M}_{\theta_0}^s, \theta_0) = \frac{J_{\theta_0}}{2}.
\]

[Assumption 1]: The map \( \theta \mapsto \rho_\theta \) is \( C^1 \) and \( \rho_\theta > 0 \).

[Assumption 2]: The map \( \theta \mapsto \text{Tr} \rho_\theta \frac{L_{\theta_0}}{J_{\theta_0}} \) is injective i.e., one-to-one.

[Construction of \( \tilde{M}_{\theta_0}^s \): We perform the POVM \( E(L_{\theta_0}/J_{\theta_0}) \) for all systems. The estimated value is determined to be the mean value plus \( \theta_0 \).]
**Proof of Lemma 14**: From Assumption 2, the weak consistency is satisfied. Let \( \delta > 0 \) be a sufficiently small number. Define the function

\[
\phi_{\theta, \theta_0}(s) := \text{Tr} \rho_{\theta} \exp \left( s \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\text{Tr} \rho_{\theta} L_{\theta_0}}{J_{\theta_0}} \right) \right).
\]  

(54)

Since \( \| L_{\theta_0} \| < \infty \) and \( \text{Tr} \rho_{\theta} \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\text{Tr} \rho_{\theta} L_{\theta_0}}{J_{\theta_0}} \right) = 0 \), we have

\[
\lim_{s \to 0} \frac{\phi_{\theta, \theta_0}(s) - 1}{s^2} = \frac{1}{2} \text{Tr} \rho_{\theta} \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\text{Tr} \rho_{\theta} L_{\theta_0}}{J_{\theta_0}} \right)^2.
\]

When \( \| \theta - \theta_0 \| \) is sufficiently small, the function \( x \to \sup_s (xs - \log \phi_{\theta, \theta_0}(s)) \) is continuous in \((-\delta, \delta)\). Using Cramér’s theorem [7], we have

\[
\lim_{n \to \infty} \frac{1}{n} \log P_{\theta_0}^{M_{s,n}} \left\{ |\hat{\theta} - \theta_0| > \varepsilon \right\} = \min \left\{ \sup_{s} (\varepsilon s - \log \phi_{\theta, \theta_0}(s)), \sup_{s'} (-\varepsilon s' - \log \phi_{\theta, \theta_0}(s')) \right\}
\]

for \( \varepsilon < \delta \). Taking the limit \( \varepsilon \to 0 \), we have

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\varepsilon^2 n} P_{\theta_0}^{M_{s,n}} \left\{ |\hat{\theta} - \theta_0| > \varepsilon \right\} = \min \left\{ \lim_{\varepsilon \to 0} \sup_{s} (\varepsilon s - \log \phi_{\theta, \theta_0}(s)), \lim_{\varepsilon \to 0} \sup_{s'} (-\varepsilon s' - \log \phi_{\theta, \theta_0}(s')) \right\} = \frac{1}{2} c_{\theta, \theta_0},
\]

where

\[
c_{\theta, \theta_0} := \text{Tr} \rho_{\theta} \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\text{Tr} \rho_{\theta} L_{\theta_0}}{J_{\theta_0}} \right)^2.
\]

because

\[
\varepsilon s - \log \phi_{\theta, \theta_0}(s) \cong \varepsilon s - \log(1 + \frac{1}{2} c_{\theta, \theta_0} s^2) \cong \varepsilon s - \frac{1}{2} c_{\theta, \theta_0} s^2 \]

\[
= -\frac{c_{\theta, \theta_0}}{2} \left( s - \frac{\varepsilon}{c_{\theta, \theta_0}} \right) + \frac{\varepsilon^2}{2 c_{\theta, \theta_0}}.
\]

The above convergence is uniform for the neighborhood of \( \theta_0 \). Taking the limit \( \theta \to \theta_0 \), we have

\[
\lim_{\theta \to \theta_0} \text{Tr} \rho_{\theta} \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\text{Tr} \rho_{\theta} L_{\theta_0}}{J_{\theta_0}} \right)^2 = J_{\theta_0}^{-1} = \text{Tr} \rho_{\theta_0} \left( \frac{L_{\theta_0}}{J_{\theta_0}} - \frac{\text{Tr} \rho_{\theta_0} L_{\theta_0}}{J_{\theta_0}} \right)^2.
\]

Thus, we can check (53) and the strong consistency in the neighborhood of \( \theta_0 \).
However, this sequence of estimators $\hat{M}^s_\delta$ depends on the true parameter $\theta_0$. We should construct a sequence of estimators that satisfies the strong consistency condition and attains the bound $\frac{J_{\theta_0}}{2}$ at all points $\theta_0$. Since such a construction is too difficult, we introduce another strong consistency condition that is weaker than the above and under which inequality (45) holds. We construct a sequence of estimators that satisfies this strong consistency condition and attains the bound given in (45) for all $\theta$ in a weak sense.

[Second strong consistency condition]: A sequence of estimators $\hat{M} = \{M^n\}$ is called second strongly consistent if there exists a sequence of functions $\{\beta_m(\hat{M}, \theta, \varepsilon)\}_{m=1}^{\infty}$ such that

\[
\begin{align*}
&\lim_{m \to \infty} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta_m(\hat{M}, \theta, \varepsilon) = \alpha(\hat{M}, \theta). \\
&\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta_m(\hat{M}, \theta, \varepsilon) \leq \alpha(\hat{M}, \theta) \text{ holds. Its LHS converges locally uniformly to } \theta. \\
&\forall m, \exists \delta > 0 \text{ s.t. } \beta(\hat{M}, \theta, \varepsilon) \geq \beta_m(\hat{M}, \theta, \varepsilon), \text{ for } \delta > \forall \varepsilon > 0.
\end{align*}
\]

Similarly to Theorem 9, we can prove inequality (45) under the second strong consistency condition.

Under these preparations, we state a theorem with respect to the attainability of the bound $J_{\theta}$.

The following theorem can be regarded as a special case of Theorem 8 of [18].

**Theorem 15:** Under Assumptions 1 and 3, the sequence of estimators $\hat{M}^s_\delta = \{M^{s,n}_\delta\}_{n=1}^{\infty}$ (defined in the following) satisfies the second strong consistency condition and the relations

\[
\alpha(\hat{M}^s_\delta, \theta) = \alpha(\hat{M}^s_\delta, \theta) = (1 - \delta) \frac{J_{\theta}}{2}. \tag{55}
\]

The sequence of estimators $\hat{M}^s_\delta$ is independent of the unknown parameter $\theta$. Every $M^{s,n}_\delta$ is an adaptive estimator and will be defined in section 8.

Its proof is given in Appendix H.

[Assumption 3]: The following set is compact.

\[
\left\{ \left( \text{Tr} \rho_\theta \left( \frac{L_\tilde{\theta}}{J_\tilde{\theta}} - \frac{\text{Tr} \rho_\theta L_\tilde{\theta}}{J_\tilde{\theta}} \right)^2 \right)^{-1}, \text{Tr} \rho_\theta \left( \frac{L_\tilde{\theta}}{J_\tilde{\theta}} - \frac{\text{Tr} \rho_\theta L_\tilde{\theta}}{J_\tilde{\theta}} \right)^2 \right\}_\forall \theta, \tilde{\theta} \in \Theta
\]

If the state family is included by a bounded closed set consisting of positive definite operators, Assumption 3 is satisfied.
[Construction of $\tilde{M}^\delta_\theta$]: We perform a faithful POVM $M_f$ (defined in the following) for the first $\delta n$ systems. Then, the data $(\omega_1, \ldots, \omega_{\delta n})$ obey the probability family $\{P^M_{\theta'} | \theta \in \Theta\}$. We denote the maximum likelihood estimator (MLE) w.r.t. the data $(\omega_1, \ldots, \omega_{\delta n})$ by $\hat{\theta}$. Next, we perform the measurement $E(L_{\hat{\theta}})$ defined by the spectral measure of $L_{\hat{\theta}}$ for other $(1 - \delta)n$ systems. Then, we have data $(\omega_{\delta n+1}, \ldots, \omega_n)$. We decide the final estimated value $T^n_{\hat{\theta}}$ as

$$\text{Tr} \rho T^n_{\hat{\theta}} L_{\hat{\theta}} = \frac{1}{(1 - \delta)n} \sum_{i=\delta n+1}^n \omega_i.$$

\textbf{Definition 16:} A POVM $M$ is called \textit{faithful}, if the map $\rho \in S(\mathcal{H}) \mapsto P^M_{\rho}$ is one-to-one.

An example of faithful POVM, which is a POVM taking values in the set of pure states on $\mathcal{H}$, is given by $M_h(d\rho) := k \rho \nu(d\rho)$, where $\nu$ is the invariant (w.r.t. the action of SU($\mathcal{H}$)) probability measure on the set of pure states on $\mathcal{H}$. As another example, if $L_1, \ldots, L_{k^2-1}$ is a basis of the space of self-adjoint traceless operators, a disjoint random combination of PVMs $E(L_1), \ldots, E(L_{k^2-1})$ is faithful. Note that a disjoint random combination is defined in section 4.

\textbf{Remark 17:} By dividing $n$ systems into $\sqrt{n}$ and $n - \sqrt{n}$ systems, Gill and Massar [11] constructed an estimator which asymptotically attains the optimal bound w.r.t. MSE, and Hayashi and Matsumoto [17] constructed a similar estimator by dividing them into $b_n$ and $n - b_n$ systems, where $\lim_{n} \frac{b_n}{n} = 0$. However, in our proof, it is difficult to show the attainability of the bound (45) in such a division. Perhaps, there may exist a family in which such an estimator does not attain the bound (45). At least, it is essential in our proof that the number of the first group $b_n$ satisfy $\lim_{n} \frac{b_n}{n} > 0$.

Conversely, as is mentioned in Theorems 13 and 18, by dividing $n$ systems into $\sqrt{n}$ and $n - \sqrt{n}$ systems, we can construct an estimator attaining the bound (34) at one point.

We must use quantum correlations in the quantum apparatus to achieve the bound $\tilde{J}^w_{\theta}$. The following theorem can be easily extended to the multi-parameter case.

\textbf{Theorem 18:} We assume Assumption 1 and that $D(\rho_{\theta'} \| \rho_{\theta_1}) < \infty$ for $\forall \theta_1, \forall \theta' \in \Theta$. Then, for any $\theta_1 \in \Theta$, the sequence of estimators $\tilde{M}^w_{\theta_1}$ =
\[ \{ M_{\theta_1}^{w,n} \}_{n=1}^{\infty} \text{satisfies the weak consistency condition (31), and the equations} \]
\[
\begin{align*}
\beta(M_{\theta_1}^{w}, \theta_1, \varepsilon) &= \beta(M_{\theta_1}^{w}, \theta_1, \varepsilon) = \inf_{\theta' \in \Theta} \{ D(\rho_{\theta'} || \rho_{\theta_1}) | | \theta_1 - \theta' | > \varepsilon \}, \\
\alpha(M_{\theta_1}^{w}, \theta_1) &= \alpha(M_{\theta_1}^{w}, \theta_1) = \frac{\bar{J}_{\theta_1}}{2}.
\end{align*}
\]

The sequence of estimators \( \bar{M}_{\theta_1} \) depends on the unknown parameter \( \theta_1 \) but not on \( \varepsilon > 0 \).

Its proof is given in Appendix I. In the following construction, \( M_{\theta_1}^{w,n} \) is constructed from the PVM \( E_{\theta_1}^{n} \), which is defined from a group-theoretical viewpoint in in Appendix J.3.

[Construction of \( M_{\theta_1}^{w,n} \):] We divide the \( n \) systems into two groups. We perform a faithful POVM \( M_f \) for the first group of \( \sqrt{n} \) systems. Then, the data \( (\omega_1, \ldots, \omega_{\sqrt{n}}) \) obey the probability \( P_{\theta}^{M_f} \). We let \( \hat{\theta} \) be the MLE of the data \( (\omega_1, \ldots, \omega_{\sqrt{n}}) \) under the probability family \( \{ P_{\theta}^{M_f} | \theta \in \Theta \} \). Next, we perform the correlational PVM \( E_{\theta_1}^{n-\sqrt{n}} \) for the composite system which consists of the other group of \( n - \sqrt{n} \) systems. Then, the data \( \omega \) obeys the probability \( P_{\theta}^{E_{\theta_1}^{n-\sqrt{n}}} \). If \( e^{(1-\delta_{n-\sqrt{n}})D(\rho_{\hat{\theta}} || \rho_{\theta_1})} P_{\theta}^{E_{\theta_1}^{n-\sqrt{n}}} (\omega) \geq P_{\theta}^{E_{\theta_1}^{n-\sqrt{n}}} (\omega) \), the estimated value \( T_n \) is decided to be \( \theta_1 \), where \( \delta_n := \frac{1}{n^{\frac{1}{2}}} \). If not, \( T_n \) is decided to be \( \hat{\theta} \).

The following lemma proven in Appendix J plays an important role in the proof of Theorem 18.

**Lemma 19:** For three parameters \( \theta_0, \theta_1 \) and \( \theta_2 \) and \( \delta > 0 \), the inequalities
\[
P_{\theta_0}^{E_{\theta_1}^{n}} \left\{ -\frac{1}{n} \log P_{\theta_2}^{E_{\theta_1}^{n}} (\omega) + \Tr \rho_{\theta_0} \log \rho_{\theta_2} \geq \delta \right\} 
\leq \exp -n \left( \sup_{0 \leq t \leq 1} (\delta - \Tr \rho_{\theta_0} \log \rho_{\theta_2})t \right)
- t \left( k + 1 \right) \log (n + 1) \frac{n}{n} - \log \Tr \rho_{\theta_0} \rho_{\theta_2} - t \right) \tag{58}
\]
\[
P_{\theta_0}^{E_{\theta_1}^{n}} \left\{ -\frac{1}{n} \log P_{\theta_1}^{E_{\theta_1}^{n}} (\omega) \right\} - \Tr \rho_{\theta_0} \log \rho_{\theta_1} \geq \delta \right\} 
\leq \exp -n \left( \sup_{0 \leq t} (\delta + \Tr \rho_{\theta_0} \log \rho_{\theta_1})t - \log \Tr \rho_{\theta_0} \rho_{\theta_1}^t \right) \tag{59}
\]
hold.
We obtain the following theorem as a review of the above discussion.

**Theorem 20:** From Theorems 5, 9 and 15 and Lemma 14, we have the equations

\[
\sup_{\mathcal{M}} \limsup_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\mathcal{M}, \theta, \varepsilon) = \sup_{\mathcal{M}} \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\mathcal{M}, \theta, \varepsilon) = \frac{\tilde{J}_\theta}{2}
\]

(60)

\[
\sup_{\mathcal{M}} \liminf_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \beta(\mathcal{M}, \theta, \varepsilon) = \tilde{J}_\theta
\]

(61)

as an operational comparison of \(\tilde{J}_\theta\) and \(J_\theta\) under Assumptions 1, 2 and 3.

We can replace \(\beta(\mathcal{M}, \theta, \varepsilon)\) with \(\beta(\mathcal{M}, \theta, \varepsilon)\) in equations (60).

We can also prove (30) as a consequence of equations (60) and (61).

### 8. Adaptive Estimators

In this section, we assume that the dimension of the Hilbert space \(\mathcal{H}\) is finite. We consider estimators whose POVM is adaptively chosen from the data. We choose the \(l\)-th POVM \(M_l(\bar{\omega}_{l-1})\) on \(\mathcal{H}\) from \(l-1\) data \(\bar{\omega}_{l-1} := (\omega_1, \ldots, \omega_{l-1})\). Its POVM \(M^n\) is described by

\[
M^n(\bar{\omega}_n) := M_1(\omega_1) \otimes M_2(\omega_1; \omega_2) \otimes \cdots \otimes M_n(\bar{\omega}_{n-1}; \omega_n).
\]

(62)

In this setting, the estimator is written as the pair \(\mathcal{E}_n = (M^n, T_n)\) of the POVM \(M^n\) satisfying (62) and the function \(T_n : \Omega^n \mapsto \Theta\). Such an estimator \(\mathcal{E}_n\) is called an adaptive estimator. As a larger class of POVMs, the separable POVM is well known. A POVM \(M^n\) on \(\mathcal{H}^{\otimes n}\) is called separable if it is written as

\[
M^n = \{M_1(\omega) \otimes \cdots \otimes M_n(\omega)\}_{\omega \in \Omega}
\]

on \(\mathcal{H}^{\otimes n}\), where \(M_i(\omega)\) is a positive semi-definite operator on \(\mathcal{H}\). For any separable estimator \((M^n, T_n)\), the relations

\[
D^{M^n}(\theta||\theta') = \sum_{\omega \in \Omega} \prod_{l=1}^n \text{Tr} \rho_\theta M_l'(\omega) \log \frac{\prod_{l=1}^n \text{Tr} \rho_\theta M_l(\omega)}{\prod_{l=1}^n \text{Tr} \rho_{\theta'} M_l(\omega)}
\]

\[
= \sum_{\omega \in \Omega} \prod_{l=1}^n \text{Tr} \rho_\theta M_l'(\omega) \sum_{l=1}^n \log \frac{\text{Tr} \rho_\theta M_l(\omega)}{\text{Tr} \rho_{\theta'} M_l(\omega)}
\]

\[
= \sum_{l=1}^n \sum_{\omega \in \Omega} a_{\theta,l}(\omega) \text{Tr} \rho_\theta M_l(\omega) \frac{a_{\theta,l}(\omega) \text{Tr} \rho_\theta M_l(\omega)}{a_{\theta,l}(\omega) \text{Tr} \rho_{\theta'} M_l(\omega)}
\]

\[
= \sum_{l=1}^n D^{M_{\theta,l}}(\theta||\theta') \leq n \sup_{M:\text{POVM on } \mathcal{H}} D^M(\theta||\theta')
\]

(63)
hold, where the POVM \( M_{\theta,l} \) on \( \mathcal{H} \) is defined by
\[
M_{\theta,l}(\omega) := a_{\theta,l}(\omega) M_l(\omega), \quad a_{\theta,l}(\omega) := \left( \prod_{l' \neq l} \text{Tr} \rho_0 M_{l'}(\omega) \right).
\]

**Theorem 21:** If a sequence of separable estimators \( \vec{M} = \{E_n\} = \{(M^n, T_n)\} \) satisfies the weak consistency condition, the inequalities
\[
\beta(\vec{M}, \theta_1, \varepsilon) \leq \inf_{|\theta - \theta_1| > \varepsilon} \sup_{\mathcal{M} : \text{POVM on } \mathcal{H}} D^M(\theta || \theta_1) \quad (64)
\]
\[
\alpha(\vec{M}, \theta_1) \leq \frac{J_{\theta_1}}{2} \quad (65)
\]
hold.

**Proof:** Similarly to (35), the monotonicity of quantum relative entropy yields
\[
-\log P_{\theta_1}^M \left\{ |T_n(\tilde{\omega}_n) - \theta_1| > \varepsilon \right\} \leq \frac{D^M_n(\theta_1 || \theta_1) + h(P_n)}{nP_n},
\]
where \( P_n := P_{\theta_1}^M \left\{ |T_n(\tilde{\omega}_n) - \theta_1| > \varepsilon \right\}. \) From the weak consistency, we have \( P_n \to 1. \) Thus, we obtain (64) from (63). Since \( \mathcal{H} \) is finite-dimensional, the set of extremal points of POVMs is compact. Therefore, the convergence \( \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} D^M(\theta_1 + \varepsilon || \theta_1) \) is uniform w.r.t. \( M. \) This implies that
\[
\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \sup_{\mathcal{M} : \text{POVM on } \mathcal{H}} D^M(\theta_1 + \varepsilon || \theta_1) = \frac{J_{\theta_1}}{2}. \quad (66)
\]
The last equation is derived from (29).

The preceding theorem holds for any adaptive estimator. As a simple extension, we can define an \( m \)-adaptive estimator that satisfies (62) when every \( M_l(\tilde{\omega}_{l-1}) \) is a POVM on \( \mathcal{H}^m. \) As a corollary of Theorem 21, we have the following.

**Corollary 22:** If a sequence of \( m \)-adaptive estimators \( \vec{M} = \{E_n\} = \{(M^n, T_n)\} \) satisfies the weak consistency condition, then the inequalities
\[
\beta(\vec{M}, \theta_1, \varepsilon) \leq \inf_{|\theta - \theta_1| > \varepsilon} \sup_{\mathcal{M} : \text{POVM on } \mathcal{H}^m} \frac{1}{m} D^M(\theta || \theta_1) \quad (67)
\]
\[
\alpha(\vec{M}, \theta_1) \leq \frac{J_{\theta_1}}{2} \quad (68)
\]
hold.
Now, we obtain the equation
\[
\lim_{m \to \infty} \lim_{\varepsilon \to 0} \sup_{\vec{M}} \frac{1}{\varepsilon^2} \beta(\vec{M}, \theta, \varepsilon) = \frac{J_0}{2}. \tag{69}
\]

The part of $\geq$ holds because an adaptive estimator attaining the bound is constructed in Theorem 15, and the part of $\leq$ follows from (67) and the equation
\[
\lim_{\varepsilon \to 0} \sup_{M: \text{POVM on } H^\otimes m} \frac{1}{\varepsilon^2 m} D^M(\theta_1 + \varepsilon \parallel \theta_1) = \frac{J_{\theta_1}}{2}, \tag{71}
\]
which is proven in a similar manner as (66).

9. Difference in Order among Limits and Supremums

Theorem 20 yields another operational comparison as
\[
\lim_{\varepsilon \to 0} \sup_{M: \text{POVM on } H^\otimes m} \frac{1}{\varepsilon^2 m} D^M(\theta_1 + \varepsilon \parallel \theta_1) = \frac{J_{\theta_1}}{2}, \tag{72}
\]
Equation (70) equals (61) and equation (71) follows from Theorem 23. Therefore, the difference between $\frac{J_0}{2}$ and $\frac{J_{\theta_1}}{2}$ can be regarded as the difference in the order of $\lim_{\varepsilon \to 0}$ and $\sup_{\vec{M}}: \text{SC}$. This comparison was naively discussed by Nagaoka [31, 32].

**Theorem 23:** We adopt Assumption 1 in Theorem 15 and $D(\rho_{\theta'} \parallel \rho_{\theta_1}) < \infty$ for $\forall \theta' \in \Theta$. For any $\delta > 0$, there exists a sequence $\vec{M}_{\theta_0}^{m, \delta} = \{M_{\theta_0}^{m, \delta, n}\}$ of $m$-adaptive estimators satisfying the strong consistency condition and the inequality
\[
\lim_{n \to \infty} \frac{-1}{nm} \log P_{\theta_0}^{M_{\theta_0}^{m, \delta, n}} \{ |\hat{\theta} - \theta_0| > \varepsilon \} \geq (1 - \delta) \inf \{ D(\theta \parallel \theta_0) | |\theta - \theta_0| > \varepsilon \} - \frac{(1 - \delta) (k - 1) \log(m + 1)}{m}.
\]

However, using Theorem 23, we obtain a stronger equation than (71):
\[
\lim_{m \to \infty} \lim_{\varepsilon \to 0} \sup_{\vec{M}} \frac{1}{\varepsilon^2} \beta(\vec{M}, \theta, \varepsilon) = \frac{J_{\theta}}{2}. \tag{72}
\]
where $m$-ASC at $\theta$ denotes $m$-adaptive and is strongly consistent at $\theta$. This equation is in contrast with (69). Of course, the part of $\leq$ for (72) follows from (67). The part of $\geq$ for (72) is derived from the above theorem.

The following two lemmas are essential for our proof of Theorem 23.

**Lemma 24:** For two parameters $\theta_1$ and $\theta_0$, the inequality

$$m D(\theta_0 \| \theta_1) - (k - 1) \log (m + 1) \leq D^m_{\theta_1}(\theta_0 \| \theta_1) \leq m D(\theta_0 \| \theta_1)$$

(73)

holds, where the PVM $E^m_{\theta_1}$ on $H^{\otimes m}$ is defined in Appendix J.3. It is independent of $\theta_0$.

This lemma was proven by Hayashi [15] and can be regarded as an improvement of Hiai and Petz’s result [22]. However, Hiai and Petz’s original version is sufficient for our proof of Theorem 23. For the reader’s convenience, the proof is presented in Appendix J.3.

**Lemma 25:** Let $Y$ be a curved exponential family and $X$ be an exponential family including $Y$. For a curved exponential family and an exponential family, see Chap 4 in Amari and Nagaoka [1] or Barndorff-Nielsen [5]. In this setting, for $n$-i.i.d. data, the MLE $T^{ML}_{X,n}(\omega^n)$ for the exponential family $X$ is a sufficient statistic for the curved exponential family $Y$, where $\omega_n := (\omega_1, \ldots, \omega_n)$. Using the map $T : X \rightarrow Y$, we can define an estimator $T \circ T^{ML}_{X,n}$, and for an estimator $T_Y$, there exists a map $T : X \rightarrow Y$ such that $T_Y = T \circ T^{ML}_{X,n}$. We can identify a map $T$ from $X$ to $Y$ with a sequence of estimators $T \circ T^{ML}_{X,n}(\omega_n)$. We define the map $T_{\theta_0} : X \rightarrow Y$ as

$$T_{\theta_0} := \arg \min_{\theta \in Y} \{D(x \| \theta) | D(\theta_0 \| \theta) \leq D(x \| \theta_0)\}.$$  

(74)

When $Y$ is an exponential family (i.e., flat), $T_{\theta_0}$ coincides with the projection to $Y$. Then, the sequence of estimators corresponding to the map $T_{\theta_0}$ satisfies the strong consistency at $\theta_0$ and the equation

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log p^n_{\theta_0} \{\|T_{\theta_0} \circ T^{ML}_{X,n}(\omega_n) - \theta_0\| > \varepsilon\} = \inf_{\theta \in Y} \{D(\theta_0 \| \theta) | \|\theta - \theta_0\| > \varepsilon\}$$

(75)

holds.

**Proof:** It is well known that for any subset $X' \subset X$, the equation

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log p^n_{\theta_0} \{T^{ML}_{X,n}(\omega_n) \in X'\} = \inf_{x \in X'} D(x \| \theta_0)$$

(76)

holds. For the reader’s convenience, we present a proof of (76) in Appendix K. Thus, equation (75) follows from (74) and (76). If $Y$ is an exponential
family, then the estimator $T_{\theta_0} \circ T_{X,n}^{ML}$ coincides with the MLE and satisfies the strong consistency. Otherwise, we choose a neighborhood $U$ of $\theta_0$ so that we can approximate the neighborhood $U$ by the tangent space. The estimator $T_{\theta_0} \circ T_{X,n}^{ML}$ can be approximated by the MLE and satisfies the strong consistency at $U$. Thus, it also satisfies the strong consistency at $\theta_0$.

Proof of Theorem 23: Let $M = \{M_i\}$ be a faithful POVM defined in section 7.2 such that the number of operators $M_i$ is finite. For any $m$ and any $\delta > 0$, we define the POVM $M_{\theta_0}^m$ to be the disjoint random combination of $M \times m$ and $E_{\theta_0}^m$ with the ratio $\delta : 1 - \delta$. Note that a disjoint random combination is defined in section 4. From the definition of $M_{\theta_0}^m$, the inequality

$$(1 - \delta)D_{\theta_0}^m (\theta \parallel \theta) \leq D_{\theta_0}^M (\theta \parallel \theta)$$

holds. Since the map $\theta \mapsto P^M_{\theta}$ is one-to-one, the map $\theta \mapsto P^m_{\theta_0}$ is also one-to-one. Since $M$ and $E_{\theta_0}^m$ are finite-resolutions of the identity, the one-parameter family $\{P^m_{\theta_0} \mid \theta \in \Theta\}$ is a subset of multi-nominal distributions $X$, which is an exponential family. Applying Lemma 25, we have

$$\lim_{n \to \infty} \frac{-1}{nm} \log P^m_{\theta_0 \times n} \{|T_{\theta_0} \circ T_{X,n}^{ML} (\bar{\omega}_n) - \theta_0| > \varepsilon\} = \frac{1}{M} \inf_{\theta \in \Theta} \{D_{\theta_0}^m (\theta \parallel \theta_0) \mid |\theta - \theta_0| > \varepsilon\}$$

$$\geq \frac{(1 - \delta)}{m} \inf \{D_{\theta_0}^m (\theta \parallel \theta_0) \mid |\theta - \theta_0| > \varepsilon\}$$

$$\geq (1 - \delta) \inf \{D(\theta \parallel \theta_0) \mid |\theta - \theta_0| > \varepsilon\} - \frac{(1 - \delta)(k - 1) \log (m + 1)}{m},$$

where the first inequality follows from (77) and the second inequality follows from (73). ■

Remark 26: In the case of the one-parameter equatorial spin 1/2 system state family, the map $\theta \mapsto P^m_{\theta_0}$ is not one-to-one. Therefore, we must treat not $E_{\theta_0}^m$ but $M_{\theta_0}^m$.

Conclusions

It has been clarified that SLD Fisher information $J_{\theta}$ gives the essential large deviation bound in the quantum estimation and KMB Fisher information $\tilde{J}_{\theta}$ gives the large deviation bound of consistent superefficient estimators. Since estimators attaining the bound $\frac{J_{\theta}}{2}$ are unnatural, the bound $\frac{\tilde{J}_{\theta}}{2}$ is
more important from the viewpoint of quantum estimation than the bound \( \frac{J}{2} \). On the other hand, as is mentioned in A, concerning a quantum analogue of information geometry from the viewpoint of e-connections, KMB is the most natural among the quantum versions of Fisher information. The interpretation of these two facts which seem to contradict each other, remains a problem. Similarly, it is a future problem to explain geometrically the relationship between the change of the orders of limits and the difference between the two quantum analogues of Fisher information.

**Acknowledgments**

The author wishes to thank Professor H. Nagaoka for encouragement and essential advice regarding this manuscript. He also wishes to thank Professor K. Matsumoto for useful advice regarding this manuscript. He is grateful to Professor S. Amari and Professor A. Tomita and two anonymous referees, whose comments helped to considerably improve the presentation.

**Appendix**

A. Brief Review of Information-geometrical Properties of \( J_\theta, \tilde{J}_\theta \) and \( \check{J}_\theta \)

The quantum analogues of Fisher information \( J_\theta, \tilde{J}_\theta \) and \( \check{J}_\theta \) are obtained from the the inner products \( J_\rho, \tilde{J}_\rho \) and \( \check{J}_\rho \) on the linear space consisting of self-adjoint operators:

\[
\tilde{J}_\rho(A, B) := \text{Tr} A\tilde{L}_B, \quad \int_0^1 \rho^t\tilde{L}_B\rho^{1-t} \, dt = B
\]

\[
J_\rho(A, B) := \text{Tr} AL_B, \quad \frac{1}{2}(L_B\rho + \rho L_B) = B
\]

\[
\check{J}_\rho(A, B) := \text{Tr} A\check{L}_B, \quad B = \rho\check{L}_B
\]

in the following way:

\[
J_\theta = J_\rho \left( \frac{d\rho_\theta}{d\theta}, \frac{d\rho_\theta}{d\theta} \right), \quad \tilde{J}_\theta = \tilde{J}_\rho \left( \frac{d\rho_\theta}{d\theta}, \frac{d\rho_\theta}{d\theta} \right), \quad \check{J}_\theta = \check{J}_\rho \left( \frac{d\rho_\theta}{d\theta}, \frac{d\rho_\theta}{d\theta} \right).
\]

In the multi-dimensional case, these are regarded as metrics as follows. For example, we can define a metrics

\[
\langle \partial_i, \partial_j \rangle = J_\rho \left( \frac{\partial \rho_\theta}{\partial \theta^i}, \frac{\partial \rho_\theta}{\partial \theta^j} \right)
\]

on the tangent space at \( \theta \), and the RHS of (78) is called SLD Fisher matrix.
In quantum setting, any information precessing is described by a trace-preserving CP (completely positive) map \( C : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}') \). These inner product satisfy the monotonicity:

\[
J_{\rho_0} \left( \frac{d\rho_0}{d\theta}, \frac{d\rho_0}{d\theta} \right) \geq J_{C(\rho_0)} \left( \frac{dC(\rho_0)}{d\theta}, \frac{dC(\rho_0)}{d\theta} \right)
\]

\[
\tilde{J}_{\rho_0} \left( \frac{d\rho_0}{d\theta}, \frac{d\rho_0}{d\theta} \right) \geq \tilde{J}_{C(\rho_0)} \left( \frac{dC(\rho_0)}{d\theta}, \frac{dC(\rho_0)}{d\theta} \right)
\]

\[
\check{J}_{\rho_0} \left( \frac{d\rho_0}{d\theta}, \frac{d\rho_0}{d\theta} \right) \geq \check{J}_{C(\rho_0)} \left( \frac{dC(\rho_0)}{d\theta}, \frac{dC(\rho_0)}{d\theta} \right)
\]

for a one-parametric density family \( \{\rho_\theta \in \mathcal{S}(\mathcal{H}) | \theta \in \Theta \subset \mathbb{R} \} \) \cite{1}. These inequalities can be regarded as the quantum versions of (5). An inner product satisfying the above is called a monotone inner product. According Petz \cite{34}, the inner product \( \check{J}_\rho \) is the maximum one among normalized monotone inner products, and the inner product \( J_\rho \) is the minimum one.

In the information geometry community, we usually discuss the torsion. As is known within this community, \( \alpha \)-connection is a generalization of \( e \)-connection. The torsion of \( \alpha \)-connection concerning Fisher inner product vanishes in any distribution family \cite{1}. In quantum setting, we can define the \( e \)-connections with respect to several quantum Fisher inner products. One may expect that in a quantum setting, its torsion vanishes in any density family. However, for only the inner product \( \check{J}_\rho \), the torsion of \( e \)-connection vanishes in any density family \cite{1}. Thus, KMB Fisher information seems the most natural quantum analogue of Fisher information, from an information-geometrical viewpoint.

**B. Proof of (15)**

From (14), we can calculate as

\[
D(\rho_{\theta+\varepsilon}||\rho_\theta) = \text{Tr} \left( \rho_{\theta+\varepsilon}(\log \rho_{\theta+\varepsilon} - \log \rho_\theta) \right)
\]

\[
\cong \text{Tr} \left( \rho_\theta + \frac{d\rho_\theta}{d\theta} \varepsilon \right) \left( \frac{d\log \rho_\theta}{d\theta} \varepsilon + \frac{1}{2} \frac{d^2 \log \rho_\theta}{d\theta^2} \varepsilon^2 \right)
\]

\[
= \text{Tr} \left( \rho_\theta \tilde{L}_\theta \right) \varepsilon + \left( \text{Tr} \left( \frac{d\rho_\theta}{d\theta} \tilde{L}_\theta \right) + \frac{1}{2} \text{Tr} \left( \rho_\theta \frac{d^2 \log \rho_\theta}{d\theta^2} \right) \right) \varepsilon^2.
\]

(79)

Next, we calculate the above coefficients

\[
\text{Tr} \left( \rho_\theta \tilde{L}_\theta \right) = \int_0^1 \text{Tr} \left( \rho_\theta \tilde{L}_\theta \rho_\theta^{-1-t} \right) dt = \text{Tr} \left( \frac{d\rho_\theta}{d\theta} \right) = 0.
\]

(80)
Using (80) and (14), we have
\[
\text{Tr} \left( \rho \frac{d^2 \log \rho}{d\theta^2} \right) = \frac{d}{d\theta} \left( \text{Tr} \left( \rho \frac{d \log \rho}{d\theta} \right) \right) - \text{Tr} \left( \frac{d\rho}{d\theta} \frac{d \log \rho}{d\theta} \right)
\]
\[
= - \text{Tr} \left( \frac{d\rho}{d\theta} L_{\theta} \right) = -\tilde{J}_{\theta}.
\] (81)

From (79), (80) and (81), we obtain
\[
D(\rho_{\theta+\epsilon} \| \rho_{\theta}) \approx \frac{1}{2} \tilde{J}_{\theta} \epsilon^2.
\]

C. Proof of Lemma 6

We define the unitary operator \( U_{\epsilon} \) as
\[
b^2(\rho_{\theta}, \rho_{\theta+\epsilon}) = 2 \left( 1 - \text{Tr} \left| \sqrt{\rho_{\theta}} \sqrt{\rho_{\theta+\epsilon}} \right| \right) = \text{Tr}(\sqrt{\rho} - \sqrt{\sigma U_{\epsilon}})(\sqrt{\rho} - \sqrt{\sigma U_{\epsilon}})^*.
\]

Letting \( W(\epsilon) \) be \( \sqrt{\rho_{\theta+\epsilon}} U_{\epsilon} \), then we have
\[
b^2(\rho_{\theta}, \rho_{\theta+\epsilon}) = \text{Tr}(W(0) - W(\epsilon))(W(0) - W(\epsilon))^*
\]
\[
\approx \text{Tr} \left( - \frac{dW}{d\epsilon}(0) \epsilon \right) \left( - \frac{dW}{d\epsilon}(0) \epsilon \right)^* \approx \text{Tr} \frac{dW}{d\epsilon}(0) \frac{dW}{d\epsilon}(0)^* \epsilon^2.
\]

As is proven in the following discussion, the SLD \( L \) satisfies
\[
\frac{dW}{d\epsilon}(0) = \frac{1}{2} LW(0).
\] (82)

Therefore, we have
\[
b^2(\rho_{\theta}, \rho_{\theta+\epsilon}) \approx \text{Tr} \frac{1}{4} LW(0) W(0)^* L \epsilon^2 = \frac{1}{4} \text{Tr} L^2 \rho_{\theta} \epsilon.
\]

We obtain (38). It is sufficient to show (82).

From the definition of the Bures distance, we have
\[
b^2(\rho_{\theta}, \rho_{\theta+\epsilon}) = \min_{U: \text{unitary}} \text{Tr}(\sqrt{\rho_{\theta}} - \sqrt{\rho_{\theta+\epsilon}} U)(\sqrt{\rho_{\theta}} - \sqrt{\rho_{\theta+\epsilon}} U)^*
\]
\[
= 2 - \max_{U: \text{unitary}} \text{Tr} \sqrt{\rho_{\theta}} \sqrt{\rho_{\theta+\epsilon}} U^* + U \sqrt{\rho_{\theta+\epsilon}} \sqrt{\rho_{\theta}}
\]
\[
= 2 - \text{Tr} \left| \sqrt{\rho_{\theta}} \sqrt{\rho_{\theta+\epsilon}} \right| + \left| \sqrt{\rho_{\theta+\epsilon}} \sqrt{\rho_{\theta}} \right|
\]
\[
= 2 - \text{Tr} \left( \sqrt{\rho_{\theta} \rho_{\theta+\epsilon} U(\epsilon)^*} + U(\epsilon) \sqrt{\rho_{\theta+\epsilon} \rho_{\theta}} \right),
\]

which implies that \( \sqrt{\rho_{\theta}} \sqrt{\rho_{\theta+\epsilon} U(\epsilon)^*} = U(\epsilon) \sqrt{\rho_{\theta+\epsilon} \rho_{\theta}} \). Therefore,
\( W(0) W(\epsilon)^* = W(\epsilon) W(0)^* \). Taking the derivative, we have
\[
W(0) \frac{dW}{d\epsilon}(0)^* = \frac{dW}{d\epsilon}(0) W(0)^*.
\]
which implies that there exists a self-adjoint operator $L$ such that

$$
\frac{dW}{d\varepsilon}(0) = \frac{1}{2}LW(0).
$$

Since $\rho_{\theta+\varepsilon} = W(\varepsilon)W(\varepsilon)^*$, we have

$$
\frac{d\rho}{d\theta}(\theta) = \frac{1}{2} \left( LW(0)W(0)^* + W(0)W(0)^*L \right).
$$

Thus, the operator $L$ coincides with the SLD.

D. Proof of (43)

Let $M = \{M_i\}$ be an arbitrary POVM. We choose the unitary $U$ satisfying

$$
U\sigma^{1/2}\rho^{1/2} = \sqrt{\rho^{1/2}\sigma\rho^{1/2}}.
$$

Using the Schwartz inequality, we have

$$
\sqrt{\mathbb{P}_\rho^M(\omega)\mathbb{P}_\sigma^M(\omega)} \\
= \sqrt{\text{Tr} \left( M_{\omega}^{1/2}\sigma^{1/2}U^* \right)^* \left( M_{\omega}^{1/2}\sigma^{1/2}U^* \right) \text{Tr} \left( M_{\omega}^{1/2}\rho^{1/2} \right)^* \left( M_{\omega}^{1/2}\rho^{1/2} \right)} \\
\geq \text{Tr} \left( M_{\omega}^{1/2}\sigma^{1/2}U^* \right)^* \left( M_{\omega}^{1/2}\rho^{1/2} \right) = \left| \text{Tr} U\sigma^{1/2}M_{\omega}\rho^{1/2} \right|.
$$

Therefore,

$$
\sum_{\omega} \sqrt{\mathbb{P}_\rho^M(\omega)\mathbb{P}_\sigma^M(\omega)} \geq \sum_{\omega} \left| \text{Tr} U\sigma^{1/2}M_{\omega}\rho^{1/2} \right| \geq \sum_{\omega} \left| \text{Tr} U\sigma^{1/2}M_{\omega}\rho^{1/2} \right| = \left| \text{Tr} U\sigma^{1/2}\rho^{1/2} \right| = \text{Tr} \sqrt{\rho^{1/2}\sigma\rho^{1/2}}.
$$

E. Proof of Lemma 7

Let $m$ and $\varepsilon$ be an arbitrary positive integer and an arbitrary positive real number, respectively. There exists a sufficiently large integer $N$ such that

$$
\frac{1}{n} \log \mathbb{P}_\theta^M \left\{ |\hat{\theta} - \theta| > \frac{\delta}{m}i \right\} \leq -\beta \left( \bar{M}, \theta, \frac{\delta}{m}i \right) + \varepsilon
$$

$$
\frac{1}{n} \log \mathbb{P}_{\theta+\delta}^M \left\{ |\hat{\theta} - (\theta + \delta)| > \frac{\delta}{m}(m-i) \right\} \leq -\beta \left( \bar{M}, \theta + \delta, \frac{\delta}{m}(m-i) \right) + \varepsilon
$$
for $i = 0, \ldots, m$ and $\forall n \geq N$. From the monotonicity (42) and the additivity (39) of quantum affinity, we perform the following evaluation:

$$-rac{n}{8}I(\rho_\theta \| \rho_{\theta+\delta}) = -\frac{1}{8}I(\rho_{\theta}^n \| \rho_{\theta+\delta}^n)$$

$$\leq \log \left( P_{\theta}^m \{ \hat{\theta} \leq \theta \} \frac{i}{2} P_{\theta+\delta}^m \{ \hat{\theta} \leq \theta \} \frac{i}{2} + P_{\theta}^m \{ \theta + \delta < \hat{\theta} \} \frac{i}{2} P_{\theta+\delta}^m \{ \theta + \delta < \hat{\theta} \} \frac{i}{2} + \sum_{i=1}^{m} P_{\theta}^m \{ \theta + \frac{\delta}{m} (i-1) < \hat{\theta} \leq \theta + \frac{\delta}{m} i \} \frac{i}{2} \times P_{\theta+\delta}^m \{ \theta + \frac{\delta}{m} (i-1) < \hat{\theta} \leq \theta + \frac{\delta}{m} i \} \frac{i}{2} \right)$$

$$\leq \log \left( P_{\theta+\delta}^m \{ |\hat{\theta} - (\theta + \delta)| \geq \delta \} \frac{i}{2} + P_{\theta}^m \{ |\hat{\theta} - \theta| > \delta \} \frac{i}{2} + \sum_{i=1}^{m} P_{\theta}^m \{ |\hat{\theta} - \theta| > \frac{\delta}{m} (i-1) \} \frac{i}{2} \times P_{\theta+\delta}^m \{ |\hat{\theta} - (\theta + \delta)| \geq \frac{\delta}{m} (m-i) \} \frac{i}{2} \right)$$

$$\leq \log \left( P_{\theta+\delta}^m \{ |\hat{\theta} - (\theta + \delta)| > \frac{\delta}{m} (m-1) \delta \} \frac{i}{2} + P_{\theta}^m \{ |\hat{\theta} - \theta| > \delta \} \frac{i}{2} + \sum_{i=1}^{m} P_{\theta}^m \{ |\hat{\theta} - \theta| > \frac{\delta}{m} (i-1) \} \frac{i}{2} \times P_{\theta+\delta}^m \{ |\hat{\theta} - (\theta + \delta)| > \frac{\delta}{m} (m-i - 1) \} \frac{i}{2} \right)$$

$$\leq \log \left( \exp \left( -\frac{n}{2} \left( \beta \left( \bar{M}, \theta, \frac{\delta}{m} (m-1) \right) - \varepsilon \right) \right) + \exp \left( -\frac{n}{2} \left( \beta \left( \bar{M}, \theta + \delta, \delta \right) - \varepsilon \right) \right) + \sum_{i=1}^{m} \exp \left( -\frac{n}{2} \left( \beta \left( \bar{M}, \theta, \frac{\delta}{m} (i-1) \right) - \varepsilon \right) \right) - \frac{n}{2} \left( \beta \left( \bar{M}, \theta + \delta, \frac{\delta}{m} (m-i - 1) \right) - \varepsilon \right) \right) \right)$$
\[
\leq \log(m + 2) \exp \left( -\frac{n}{2} \min_{0 \leq i \leq m} \left( \beta \left( \frac{\bar{M}, \theta, \delta}{m} (i - 1) \right) + \frac{\beta}{m} \left( \bar{M}, \theta + \delta, \frac{\delta}{m} (m - i - 1) - 2\varepsilon \right) \right) \right)
\]

\[
= \log(m + 2) - \frac{n}{2} \left( \min_{0 \leq i \leq m} \frac{\beta}{m} \left( \frac{\bar{M}, \theta + \delta, \frac{\delta}{m} (m - i - 1) - 2\varepsilon}{\beta \left( \frac{\bar{M}, \theta, \delta}{m} (i - 1) \right)} \right) \right),
\]

where we assume that \( \beta(\bar{M}, \theta, a) = 0 \) for any negative real number \( a \). Taking the limit \( n \to \infty \) after dividing by \( n \), we have

\[
\frac{1}{8} I(\rho_\theta \| \rho_{\theta + \delta}) \geq \frac{1}{2} \min_{0 \leq i \leq m} \left( \beta \left( \frac{\bar{M}, \theta, \delta}{m} (i - 1) \right) + \frac{\beta}{m} \left( \frac{\bar{M}, \theta + \delta, \frac{\delta}{m} (m - i - 1) - 2\varepsilon}{\beta \left( \frac{\bar{M}, \theta, \delta}{m} (i - 1) \right)} \right) \right).
\]

Since \( \varepsilon > 0 \) is arbitrary, the inequality

\[
\frac{1}{8} I(\rho_\theta \| \rho_{\theta + \delta}) \geq \frac{1}{2} \min_{0 \leq i \leq m} \left( \beta \left( \frac{\bar{M}, \theta, \delta}{m} (i - 1) \right) + \frac{\beta}{m} \left( \frac{\bar{M}, \theta + \delta, \frac{\delta}{m} (m - i - 1) - 2\varepsilon}{\beta \left( \frac{\bar{M}, \theta, \delta}{m} (i - 1) \right)} \right) \right)
\]

holds. Taking the limit \( m \to \infty \), we obtain (44).

K. Large Deviation Theory for an Exponential Family

In this section, we review the large deviation theory for an exponential family. A \( d \)-dimensional probability family is called an exponential family if there exist linearly independent real-valued random variables \( F_1, \ldots, F_d \) and a probability distribution \( p \) on the probability space \( \Omega \) such that the family consists of the probability distribution

\[
p_\theta(\omega) := \exp \left( \sum_{i=1}^{d} \theta_i F_i(\omega) - \psi(\theta) \right) p(\omega)
\]

\[
\psi(\theta) := \log \int_{\Omega} \exp \left( \sum_{i=1}^{d} \theta_i F_i(\omega) \right) p(\omega).
\]

In this family, the parametric space is given by \( \Theta := \{ \theta \in \mathbb{R}^d | 0 < \psi(\theta) < \infty \} \), the parameter \( \theta \) is called the natural parameter and the function \( \psi(\theta) \)
is called the potential. We define the dual potential \( \phi(\theta) \) and the dual parameter \( \eta(\theta) \), called the expectation parameter, as
\[
\eta_i(\theta) := \frac{\partial \psi(\theta)}{\partial \theta^i} = \log \int_{\Omega} F_i(\omega)p_{\theta}(d\omega)
\]
\[
\phi(\theta) := \max_{\theta'} \left( \sum_{i=1}^{d} \theta'^i \eta_i(\theta) - \psi(\theta') \right).
\]

From (83), we have
\[
\phi(\theta) = \sum_{i=1}^{d} \theta^i \eta_i(\theta) - \psi(\theta).
\]

In this family, the sufficient statistics are given by \( F_1(\omega), \ldots, F_d(\omega) \). The MLE \( \hat{\theta}(\omega) \) is given by \( \eta_i(\hat{\theta}(\omega)) = F_i(\omega) \). The KL divergence \( D(\theta||\theta_0) := D(p_\theta||p_{\theta_0}) \) is calculated by
\[
D(\theta||\theta_0) = \int_{\Omega} \log \left( \frac{p_\theta(\omega)}{p_{\theta_0}(\omega)} \right) p_\theta(d\omega)
\]
\[
= \int_{\Omega} \sum_{i} (\theta^i - \theta_0^i) F_i(\omega) + \psi(\theta_0) - \psi(\theta)p_\theta(d\omega)
\]
\[
= \sum_{i} (\theta^i - \theta_0^i) \eta_i(\omega) + \psi(\theta_0) - \psi(\theta) = \phi(\theta) + \psi(\theta_0) - \sum_{i} \theta_0^i \eta_i(\omega)
\]
\[
= \max_{\theta'} \left( \sum_{i} \theta'^i \eta_i(\theta) - \psi(\theta') \right) + \psi(\theta_0) - \sum_{i} \theta_0^i \eta_i(\theta)
\]
\[
= \max_{\theta'} \sum_{i} (\theta'^i - \theta_0^i) \eta_i(\theta) - \log \int_{\Omega} \exp \left( \sum_{i} (\theta'^i - \theta_0^i) F_i(\omega) \right) p_\theta(d\omega).
\]

Next, we discuss the \( n \)-i.i.d. extension of the family \( \{p_\theta|\theta \in \Theta\} \). For the data \( \omega_n := (\omega_1, \ldots, \omega_n) \in \Omega^n \), the probability distribution \( p^n_\theta(\omega_n) := p_{\theta}(\omega_1) \ldots p_{\theta}(\omega_n) \) is given by
\[
p^n_\theta(\omega_n) = \exp \left( n \sum_{i} \theta^i F_n,i(\omega_n) - n \psi(\theta) \right) p^n(d\omega_n)
\]
\[
p^n(d\omega_n) := p(d\omega_1) \ldots p(d\omega_n), \quad F_{n,i}(\omega_n) := \frac{1}{n} \sum_{k=1}^{n} F_i(\omega_k).
\]

Since the expectation parameter of the probability family \( \{p^n_\theta|\theta \in \Theta\} \) is given by \( n \eta_i(\theta) \), the MLE \( \hat{n} \eta_i(\omega_n) \) is given by
\[
n \eta_i(\hat{n} \eta_i(\omega_n)) = n F_{n,i}(\omega_n).
\]
Applying Cramér’s Theorem [7] to the random variables $F_1, \ldots, F_d$ and the distribution $p_{\theta_0}$, for any subset $S \subset \mathbb{R}^d$ we have

$$\inf_{\eta \in S} \sup_{\theta' \in \mathbb{R}^d} \left( \sum_i \theta'^i (\eta_i - E_{\theta_0}(F_i)) - \psi_{\theta_0}(\theta') \right) \leq \lim_{n \to \infty} -\frac{1}{n} \log p^n_{\theta_0} \{ \tilde{F}_n \in S \}$$

$$\leq \inf_{\eta \in \text{int} S} \sup_{\theta' \in \mathbb{R}^d} \left( \sum_i \theta'^i (\eta_i - E_{\theta_0}(F_i)) - \psi_{\theta_0}(\theta') \right),$$

where

$$E_{\theta_0}(F_i) := \int_{\Omega} F_i(\omega)p_\theta(d\omega), \quad \psi_{\theta_0}(\theta) := \int_{\Omega} \exp \left( \sum_i \theta F_i(\omega) \right) p_\theta(d\omega)$$

$$\tilde{F}_n(\tilde{\omega}_n) := (F_{n,1}(\tilde{\omega}_n), \ldots, F_{n,d}(\tilde{\omega}_n)),$$

and int$S$ denotes the interior of $S$, which is consistent with $(S^c)^c$. Since

$$\sup_{\theta' \in \mathbb{R}^d} \left( \sum_i \theta'^i (\eta_i - E_{\theta_0}(F_i)) - \psi_{\theta_0}(\theta') \right)$$

$$= \sup_{\theta' \in \mathbb{R}^d} \left( \sum_i \theta'^i (\eta_i - \eta_i(\theta_0)) - \psi(\theta') \right) + \psi(\theta_0) = D(\theta || \theta_0)$$

and the map $\theta \mapsto D(\theta || \theta_0)$ is continuous, it follows from (84) that

$$\lim_{n \to \infty} -\frac{1}{n} \log p^n_{\theta_0} \{ \hat{\theta}_n \in \Theta' \} = \inf_{\theta \in \Theta'} D(\theta || \theta_0)$$

for any subset $\Theta' \subset \Theta$, which is equivalent to (76). Conversely, if an estimator $\{T_n(\tilde{\omega}_n)\}$ satisfies the weak consistency

$$\lim_{n \to \infty} p^n_\theta \{ ||T_n(\tilde{\omega}_n) - \theta|| > \varepsilon \} \to 0, \quad \forall \varepsilon > 0, \forall \theta \in \Theta,$$

then, similarly to (33), we can prove

$$\lim_{n \to \infty} -\frac{1}{n} \log p^n_{\theta_0} \{ T_n(\tilde{\omega}_n) \in \Theta' \} \leq \inf_{\theta \in \Theta'} D(\theta || \theta_0).$$

Therefore, we can conclude that the MLE is optimal in the large deviation sense for exponential families.

References


CHAPTER 29

Estimating the Spectrum of a Density Operator

Michael Keyl and Reinhard F. Werner

Abstract. Given $N$ quantum systems prepared according to the same density operator $\rho$, we propose a measurement on the $N$-fold system that approximately yields the spectrum of $\rho$. The projections of the proposed observable decompose the Hilbert space according to the irreducible representations of the permutations on $N$ points, and are labeled by Young frames, whose relative row lengths estimate the eigenvalues of $\rho$ in decreasing order. We show convergence of these estimates in the limit $N \to \infty$, and that the probability for errors decreases exponentially with a rate we compute explicitly.

1. Introduction

The density operator of a quantum system describes the preparation of the system in all details relevant to statistical experiments. Like a classical probability distribution it cannot be measured on a single system, but can be estimated only on an ensemble sequence of identically prepared systems. In fact, if we could determine the density operator (or, in the pure case, the wave function) on a single quantum system, we could combine the measurement with a device repreparing several systems with the measured density operator, in contradiction to the well-known no-cloning theorem [22]. This points to a close connection between the problem of estimating the density operator and approximate cloning. In the case of inputs promised to be in a pure state the optimal solutions to both problems are known [20,11,2], and it turns out that in a sense the limit of the cloning problem for output number $M \to \infty$ is equivalent to the estimation problem. The “optimal” cloning transformation was shown in this case to be quite insensitive to the figure of merit defining optimality [11].

In the case of mixed input states much less is known about the cloning problem. It is likely that in this case there may be different natural figures

of merit leading to inequivalent “optimal” solutions. Even the classical version of the problem is not trivial, and is related to the so-called bootstrap technique [6] in classical statistics.

The estimation problem certainly has many solutions. In fact, any procedure of determining the density matrix through the measurement of the expectations of a suitable “quorum” of observables [18], such as in quantum state tomography [13] is a solution. Other methods include adaptive schemes [8] where the result of one measurement is used to select the next one. In all these cases, the estimate amounts to the measurement of an observable on the full input state $\rho \otimes N$, which factorizes into one-site observables. What we are concerned with here, as in the work of Vidal et al. [17], is the search for improved estimates, admitting arbitrary observables on the $N$-fold input system, including “entangled” ones. In contrast to [17], however, we are not interested in estimators that are optimal for a more or less general figure of merit, but in the asymptotic behavior if the number $N$ of input systems goes to infinity (in this context see also the work of Gill and Massar [9]).

When $\mathcal{H} \cong \mathbb{C}^d$ is the Hilbert space of a single system, the overall input density operator of the estimation problem is $\rho \otimes n$, which exists on the $N$th tensor power $\mathcal{H} \otimes N$. This space has a natural orthogonal decomposition according to the irreducible representations of the permutation group of $N$ points, acting as the permutations of the tensor factors. Equivalently, this is the decomposition according to the irreducible representations of the unitary group on $\mathcal{H}$ (see below). It is well known that this orthogonal decomposition is labeled by Young frames, i.e., by the arrangements of $N$ boxes into $d$ rows of lengths $Y_1 \geq Y_2 \geq \cdots \geq Y_d \geq 0$ with $\sum Y_\alpha = N$. There is a striking similarity here with the spectra we want to estimate, which are given by sequences of the eigenvalues of $\rho$, say, $r_1 \geq r_2 \geq \cdots \geq r_d \geq 0$, with $\sum r_\alpha = 1$. The basic idea of this paper is to show that this is not a superficial similarity: measuring the Young frame (by an observable whose eigenprojections are the projections in the orthogonal decomposition) is, in fact, a good estimate of the spectrum. More precisely, we show that the probability for the error $|Y_\alpha / N - r_\alpha|$ to be larger than a fixed $\varepsilon$ for some $\alpha$ decreases exponentially as $N \rightarrow \infty$.

The group theoretic ideas just sketched are nothing new but go back to Weyl [21] and are in the meantime a standard tool within quantum mechanics. Examples of works where similar methods are used in quantum information are [4, 17, 1, 10, 12]. In particular, [12] is closely related to the present paper because similar techniques are used there. This concerns in
particular the theory of large deviations [7], and a result by Duffield [5] on the large deviation properties of tensor powers of group presentations. This will allow us to compute the rate of exponential convergence explicitly.

2. Statement on the Result

In order to state our result, explicitly, we need to recall the decomposition theory for \( N \)-fold tensor products. Throughout, the one-particle space \( \mathcal{H} \) will be the \( d \)-dimensional Hilbert space \( \mathbb{C}^d \), with \( d < \infty \). Two group representations play a crucial role: first, the representation \( X \mapsto X \otimes N \) of the general linear group \( \text{GL}(d, \mathbb{C}) \) and, secondly, the representation \( p \mapsto S_p \) of the permutations \( p \in S_N \) on \( N \) points, represented by permuting the tensor factors:

\[
S_p \psi_1 \otimes \cdots \otimes \psi_N = \psi_{p^{-1}1} \otimes \cdots \otimes \psi_{p^{-1}N}.
\] (1)

The basic result [15] is that these two representations are “commutants” of each other, i.e., any operator on \( \mathcal{H} \otimes N \) commuting with all \( X \otimes N \) is a linear combination of the \( S_p \), and conversely. This leads to the decomposition

\[
\mathcal{H} \otimes N \cong \bigoplus_Y \mathcal{R}_Y \otimes S_Y,
\] (2)

\[
X \otimes N \cong \bigoplus_Y \pi_Y(X) \otimes 1,
\] (3)

\[
S_p \cong \bigoplus_Y 1 \otimes \hat{\pi}_Y(p),
\] (4)

where \( \pi_Y : \text{GL}(d, \mathbb{C}) \to \mathcal{B}(\mathcal{R}_Y) \) and \( \hat{\pi}_Y : S_N \to \mathcal{B}(S_Y) \) are irreducible representations, and the restriction of \( \pi_Y \) to unitary operators is unitary. The summation index \( Y \) runs over all Young frames with \( d \) rows and \( N \) boxes, as described in the Introduction. We denote by \( P_Y \) the projection onto the corresponding summand in the above decomposition.

Let us consider now the estimation problem. As already discussed in the Introduction, we are searching for an observable \( E_N \) describing a measurement on \( N \) \( d \)-level systems, whose readouts are possible spectra of \( d \)-level density operators. The set of possible spectra will be denoted by

\[
\Sigma = \left\{ s \in \mathbb{R}^d \bigg| x \triangleright 0, \sum_{j=1}^d x_j = 1 \right\},
\] (5)

where \( x \triangleright 0 \) denotes the ordering relation on \( \mathbb{R}^d \) given by

\[
s \triangleright 0 :\iff s_j > s_{j+1} \text{ for all } j = 1, \ldots, d - 1.
\] (6)
Technically, $E_N$ must be a positive operator-valued measure on this set, assigning to each measurable subset $\Delta \subset \Sigma$ a positive operator $E_N(\Delta) \in \mathcal{B}(\mathcal{H}^\otimes N)$, whose expectation in any given state is interpreted as the probability for the measurement to yield a result $s \in \Delta$.

The criterion for a good estimator $E_N$ is that, for any one-particle density operator $\rho$, the value measured on a state $\rho^\otimes N$ is likely to be close to the true spectrum $r \in \Sigma$ of $\rho$, i.e., that the probability

$$K_N(\Delta) := \text{Tr} [E_N(\Delta) \rho^\otimes N] \quad (7)$$

is small when $\Delta$ is the complement of a small ball around $r$. Of course, we will look at this problem for large $N$. So our task is to find a whole sequence of observables $E_N$, $N=1,2,\ldots$, making error probabilities like Eq. (7) go to zero as $N \to \infty$.

The search for efficient estimation strategies $E_N$ can be simplified greatly by symmetry arguments. To see this, consider a permutation $p \in S_N$. If we insert the transformed estimator $S_p E_N(\Delta) S_p^*$ into Eq. (7) we see immediately that $K_N(\Delta)$ remains unchanged. Replacing $E_N(\Delta)$ by the average $N!^{-1} \sum_{p \in S_N} S_p E_N(\Delta) S_p^*$ shows that we may assume $[E_N(\Delta), S_p] = 0$ for all permutations $p$, without loss of estimation quality. A similar argument together with the fact that the quality of the estimate is judged by some criterion not depending on the choice of a basis in $\mathcal{H}$ shows that we may assume in addition that $E_N(\Delta)$ commutes with all unitaries $U^\otimes N$. But this implies according to Eqs. (3) and (4) that $E_N(\Delta)$ must be a function of the projection operators $P_Y : \mathcal{H}^\otimes N \to \mathcal{R}_Y \otimes S_Y$ defined at the beginning of this section. If we require in addition that each $E_N(\Delta)$ be a projection, which is suggestive for ruling out unnecessary fuzziness, $E_N$ must be of the form

$$E_N(\Delta) = \sum_{Y : s_N(Y) \in \Delta} P_Y, \quad (8)$$

where $s_N$ is an arbitrary mapping assigning to each Young frame $Y$ (with $d$ rows and $N$ boxes) an estimate $s_N(Y) \in \Sigma$. In other words, the estimation proceeds by first measuring the Young frame projections $P_Y$ and then computing an estimate $s_N(Y)$ on the basis of the result $Y$.

The simplest choice is clearly to take the normalized Young frames themselves as the estimate, i.e.,

$$s_N(Y) = Y/N. \quad (9)$$

It turns out somewhat surprisingly that with this choice the $E_N(\Delta)$ from Eq. (8) form an asymptotically exact estimator. By this we mean that, for
Fig. 1. Probability distribution $\text{Tr}(\rho^{\otimes N} P_Y)$ for $d = 3$, $N = 20, 100, 500$, and $r = (0.6, 0.3, 0.1)$. The set $\Sigma$ is the triangle with corners $A = (1, 0, 0)$, $B = (1/2, 1/2, 0)$, $C = (1/3, 1/3, 1/3)$.

every $\rho$, the probability measures $K_N$ from Eq. (7) converge weakly to the point measure at the spectrum $r$ of $\rho$. Explicitly, for each continuous function $f$ on $\Sigma$ we have

$$\lim_{N \to \infty} \int_{\Sigma} f(s) K_N(ds) = \lim_{N \to \infty} \sum_Y f \left( \frac{Y}{N} \right) \text{Tr}(\rho^{\otimes N} P_Y) = f(r). \quad (10)$$

We illustrate this in Fig. 1, for $d = 3$ and $\rho$ a density operator with spectrum $r = (0.6, 0.3, 0.1)$. Then $\Sigma$ is a triangle with corners $A = (1, 0, 0)$, $B = (1/2, 1/2, 0)$, and $C = (1/3, 1/3, 1/3)$, and we plot the probabilities $\text{Tr}(\rho^{\otimes N} P_Y)$ over $Y/N \in \Sigma$. The explicit computation uses the Weyl character formula (Chap. 9, Sec. 9.1 [15]), which we do not need elsewhere in the paper.

Clearly, the distribution is peaked at the true spectrum and our claim
is that this will become exact in the limit \( N \to \infty \). To prove convergence we will use *large deviation methods* which give us not only the convergence just stated but an *exponential error estimate* of the form

\[
K_N(\Delta) \approx \exp(-N \inf_{s \in \Delta} I(s)),
\]

where \( I \) denotes a positive function on \( \Sigma \), called the *rate function*, which vanishes only for \( s = r \).

For the statement of the main theorem we say that a measurable subset \( \Delta \subset \Sigma \) has "small boundary" if its interior is dense in its closure. A typical choice for \( \Delta \) is the complement of a ball around the true spectrum.

**Theorem 1:** The estimator defined in Eqs. (8) and (9) is asymptotically exact. Moreover, we have the error estimate

\[
\lim_{N \to \infty} \frac{1}{N} \ln K_N(\Delta) = - \inf_{s \in \Delta} I(s)
\]

for any set \( \Delta \subset \Sigma \) with small boundary, where the rate function \( I : \Sigma \to [0, \infty] \) is

\[
I(s) = \sum_j s_j (\ln s_j - \ln r_j).
\]

The expression for \( I \) is the relative entropy [14] of the probability vectors \( s \) and \( r \). Relative entropies occur also as the rate functions in large deviation properties of independent identically distributed (classical [3] or quantum [19]) random variables, although there seems to be no direct way to reduce the above theorem to these standard setups.

### 3. Sketch of Proof

Rather than giving a proof of every detail, our aim here is to explain why the scaled Young frames \( Y/N \) appear in the estimation problem. The crucial observation is that the Young frame \( (Y_1, \ldots, Y_d) \) is the *highest weight* of the representation \( \pi_Y \) in the ordering \( \succ \) and this ordering is directly related to picking out the fastest growing exponential in certain integrals of the measurement \( K_N \).

The integrals we need to study are the Laplace transforms of the measures \( K_N \). We introduce the "*scaled cumulant generating function*"

\[
c(\eta) = \lim_{N \to \infty} \ln \frac{1}{N} \int_{\Sigma} K_N(ds) e^{N\eta \cdot s},
\]
where \( \eta \in \mathbb{R}^d \), and \( \eta \cdot s \) is the scalar product. If the measures \( K_N \) behave like Eq. (11) the integrand near \( s \) behaves like \( \exp N[\eta \cdot s - I(s)] \), and the largest contribution comes from the fastest growing exponential:

\[
c(\eta) = \sup_{s}[\eta \cdot s - I(s)].
\]

(15)

This is an instance of Varadhan’s theorem [16], which has a converse, the Gärtner-Ellis theorem (Theorem II.6.1 [7]): if the limit (14) exists and is differentiable then the estimate in the theorem holds, with the rate function determined from Eq. (15) by inverse Legendre transformation. We will follow Duffield [5] by computing the limit (14) from group theoretical data.

Consider the “maximally Abelian subgroup” \( C \subset \text{GL}(d, \mathbb{C}) \) of diagonal matrices

\[
\rho_h = \text{diag}[\exp(h_1), \ldots, \exp(h_d)]
\]

(16)

for \( h \in \mathbb{C}^d \). Since these commute, all the operators \( \pi_Y(\rho_h) \) commute in every representation \( \pi_Y \), and can hence be simultaneously diagonalized. The vectors \( \mu = (\mu_1, \ldots, \mu_d) \) such that \( \pi_Y(\rho_h)\psi = \exp(\mu \cdot h)\psi \) for some nonzero vector \( \psi \) are called weights of the representation \( \pi_Y \). The dimension \( m(\mu) \) of the corresponding eigenspace is called multiplicity of \( \mu \). One particular weight (with multiplicity 1) is the Young frame \( Y \) itself (interpreted as an element of \( \mathbb{R}^d \)) and it turns out that \( Y \) is the maximum (the “highest weight”) among all weights of \( \pi_Y \), in the \( \triangleright \) ordering from Eq. (6). Representation theory of semisimple Lie algebras [15] shows that each irreducible, analytic representation of \( \text{GL}(d, \mathbb{C}) \) is uniquely characterized (up to unitary equivalence) by its highest weight \( Y \).

In order to estimate the integral (14), we need the quantities \( \text{tr}(\rho^{\otimes N} P_Y) \). For simplicity we assume that \( \rho \) is nonsingular, i.e., an element of \( \text{GL}(d, \mathbb{C}) \). By Eq. (3) we have

\[
\text{Tr}(\rho^{\otimes N} P_Y) = \text{Tr} [\pi_Y(\rho) \otimes 1] = \chi_Y(\rho)\dim(S_Y),
\]

(17)

where

\[
\chi_Y(\rho) := \text{tr}[\pi_Y(\rho)]
\]

(18)

is the character of the representation \( \pi_Y \). Since \( \chi_Y(\rho) \) is unitarily invariant \( [\chi_Y(U\rho U^*) = \chi_Y(\rho)] \) we may assume that without loss of generality that \( \rho \) is diagonal and its matrix elements are arranged in descending order. Using the notation from Eq. (16) this assumption reads

\[
\rho = \rho_h \in C \text{ with } h \triangleright 0 \text{ and } \sum_j \exp(h_j) = 1.
\]

(19)
Hence we can express \( \chi_Y(\rho) \) in terms of the weights of \( \pi_Y \):

\[
\chi_Y(\rho) = \sum_\mu m(\mu) \exp(\mu \cdot h),
\]

(20)

where the sum is taken over all weights \( \mu \) of \( \pi_Y \). Since \( h \gg 0 \) and \( Y \gg \mu \) for all \( \mu \) we see that \( \exp(Y \cdot h) \) is the largest exponential. We therefore estimate

\[
\exp(Y \cdot h) \leq \chi_Y(\rho) \leq \dim(\mathcal{R}_Y) \exp(Y \cdot h). \tag{21}
\]

Hence, if we introduce for any \( h, \eta \in \mathbb{R}^d \), \( h, \eta \gg 0 \) the two expressions

\[
J(h, \eta) = \int_{\Sigma} K_N(ds) e^{N\eta \cdot s} \tag{22}
\]

\[
= \sum_Y \text{Tr}(\rho^N \pi_Y) e^{N\eta \cdot Y/N} \tag{23}
\]

\[
= \sum_Y \chi_Y(\rho) e^{\eta \cdot Y} \dim(\mathcal{S}_Y) \tag{24}
\]

and

\[
J'(h, \eta) = \sum_Y e^{(h + \eta) \cdot Y} \dim(\mathcal{S}_Y) \tag{25}
\]

we get

\[
J'(h, \eta) \leq J(h, \eta) \leq \dim(\mathcal{R}_Y) J'(h, \eta). \tag{26}
\]

If we combine this with the consequence of Weyl’s dimension formula that \( \dim(\mathcal{R}_Y) \) is bounded above by a polynomial \( p(N) \) in \( N \), uniformly in \( Y \) (Lemma 2.2 [5]), and take logarithms we get

\[
\ln J'(h, \eta) \leq \ln J(h, \eta) \leq \text{const} \times \ln N + \ln J'(h, \eta). \tag{27}
\]

Since \( J'(h, \eta) \) grows exponentially in \( N \) its logarithm is linear in \( N \) and we see that \( J(h, \eta) \) and \( J'(h, \eta) \) are asymptotically equivalent in the sense that

\[
(1/N)[\ln J(h, \eta) - \ln J'(h, \eta)] \to 0. \tag{28}
\]

In the same sense we can continue the chain of equivalences

\[
J(h, \eta) \approx J'(h, \eta) = J'(h + \eta, 0) \approx J(h + \eta, 0) \approx \int_{\Sigma} K_N(ds). \tag{29}
\]

Here we have Eq. (22) for \( J(h + \eta, 0) \), and the \( h + \eta \) dependence is contained in \( K_N(ds) \) via \( \rho_{h+\eta} \). Together with the definition of \( K_N \) in Eq. (7) this
implies
\[ J(h, \eta) \approx \int_{\Sigma} K_N(ds) = K_N(\Sigma) = \text{Tr} \left( E_N(\Sigma) \rho_{h+\eta}^{\otimes N} \right) \]  
\[ = \text{Tr} \left( \rho_{h+\eta}^{\otimes N} \right) = (\text{tr} \rho_{h+\eta})^N. \] 

Hence, if \( r_{\alpha} = \exp(h_{\alpha}) \) are the eigenvalues of a nonsingular density operator, we get for Eq. (14) the expression
\[ c(\eta) = \ln \sum_{\alpha} r_{\alpha} \exp(\eta_{\alpha}). \]

It is then a simple calculus exercise to verify the above rate function as the Legendre transform \( I(s) = \sup_{\eta} [\eta \cdot s - c(\eta)] \).

This concludes our sketch of proof. In order to expand it into a full proof, one needs to extend the computation of \( c(\eta) \) to \( \eta \neq 0 \), and prove that this extension has the required regularity properties for the application of the converse Varadhan’s theorem cited above. This has been carried out by Duffield [5] in a context that is, on the one hand wider, because it includes tensor powers of much more general representations of semisimple Lie groups, but on the other hand narrower, because it contains only the case \( \rho = d^{-1}1 \) of our theorem. However, one can extend Duffield’s result by multiplying his measures \( K_N \) by the factor \( \chi_Y(\rho)/\chi_Y(1) \) and using for this factor the estimate (21).

4. Discussion

Although the estimate we discuss is asymptotically exact, it is not at all clear whether and in what sense it might be optimal, even for finite \( N \). We have experimented with various figures of merit for estimation and found different “optimal” estimators for low \( N \), rarely coinciding with the \( E_N \) determined by Eq. (9). It is also not at all clear how much could be gained by optimization here.

An interesting extension will also be the construction of estimators for the full density operator. It is very suggestive to compose this out of the above estimator for the spectrum, and to use for each Young frame a covariant observables to estimate the eigenbasis of \( \rho \). The density of the covariant observable might be based on the highest weight vector of \( \pi_Y \).

Acknowledgements

Funding by the European Union project EQUIP (Contract No. IST-1999-11053) and financial support from the DFG (Bonn) are gratefully acknowledged.
References

PART VI

Further Topics on Quantum Statistical Inference

Chap. 31: V. Bužek, R. Derka, and S. Massar
“Optimal quantum clocks” .......................... 477

Chap. 32: A. Fujiwara
“Quantum channel identification problem” ................. 487

Chap. 33: G.M. D’Ariano
“Homodyning as universal detection” .................... 494

Chap. 34: D.F.V. James, P.G. Kwiat, W.J. Munro, and A.G. White
“On the measurement of qubits” ............................. 509
CHAPTER 30

Introduction to Part VI

1. Estimation of Quantum State-Evolution Process

The purpose of quantum statistical inference is not limited to the estimation of an unknown state, but quantum statistical inference can be applied to estimation of an unknown quantum state-evolution process. For this type of estimation, we assume that we are able to choose suitable input states as well as quantum measurements. If there is no noise, the state-evolution process is described as an action of a unitary matrix. Bužek, Derka, and Massar [Chap. 31] focused on the estimation of a one-parameter unknown unitary action

\[
\begin{pmatrix}
  e^{i\theta/2} & 0 \\
  0 & e^{-i\theta/2}
\end{pmatrix}
\]

that can operate on \( n \) systems simultaneously. Using Helstrom and Holevo’s formula [IV-1, IV-3], they showed that the optimal estimation error \( \sin^2 \frac{\theta}{2} - \hat{\theta}^2 \) goes to 0 with the rate \( \frac{\pi^2}{4n^2} \). Their optimal input state is entangled between all identical input systems. Fujiwara [VI-1] focused on that of a single unknown SU(2) action, and analyzed the merit of the input state entangled with the reference system. This problem was extended by Ballester [VI-2]. Acín et al. [VI-3] also discussed related topics. Bagan et al. [VI-4] derived the optimal estimation of the unknown \( n \)-identical SU(2) operations with use of entanglement with the reference system. They also showed that the optimal error goes to 0 with the rate of \( \frac{\pi^2}{n^2} \). They used effectively applied the Clebsch-Gordan coefficients method to this problem. Hayashi [VI-5] also treated the same problem by a different method. He derived a notable relation between this problem and that of Bužek et al., and applied the obtained relation to this problem. Furthermore, Bagan et al. [VI-6], and Chiribella et al. [VI-7] independently showed that the estimation error goes to 0 with the same rate. These papers pointed out that the multiplicity of the same irreducible representations can be regarded as the reference system, i.e., the effect of ‘self-entanglement’.

\[
\begin{align*}
\text{\footnotesize*} & \text{They proved that the error } 6 - 2\chi_1(U\hat{U}^{-1}) \text{ goes to 0 with the rate } \frac{8\pi^2}{n^2}, \text{ where } \chi_j(g) \\
& \text{denotes the trace of } g \text{ in the spin-j representation. If we focus on the error } \frac{1 - \chi_{1/2}(U\hat{U}^{-1})}{4}, \text{ the error goes to 0 with the rate } \frac{4\pi^2}{n^2} \text{ because } \chi_{1/2}(g)^2 = 1 + \chi_1(g). 
\end{align*}
\]
Moreover, Fujiwara [Chap. 32] focused on the estimation of a depolarizing channel, which is a typical example of noisy quantum state evolutions. By the careful discussion of Fisher information, he proved the advantage of the input state entangled with a reference system. He also discussed a slightly different setting. After this result, Sasaki et al. [VI-8] discussed a similar estimation problem with the Bayesian approach in a non-asymptotic setting. Fischer et al. [VI-9] focused the use of the maximally entangled input state for the estimation of the Pauli channel. Fujiwara and Imai [VI-10] and Hayashi† proved that the maximally entangled input state yields the optimal estimation of the unknown \( n \)-identical Pauli channel operations, by different methods. Furthermore, Fujiwara and Imai [VI-10] proved that only a maximally entangled input state realizes the optimal estimation in this problem. Based on this result, Fujiwara [VI-11] and Tanaka‡ treated the estimation problem of the amplitude damping channel, independently. Especially, Fujiwara [VI-11] proceeded to the estimation problem of the generalized amplitude damping channel, which is the more general and difficult part. De Martini et al. [VI-12] implemented an experiment for the estimation of an unknown unitary.

2. Tomography

The previous parts treat the estimation with the choice of the quantum measurement. In the following, we treat the estimation with a fixed quantum measurement that is experimentally available.

This type of approach is called tomography, and the infinite-dimensional case was studied first. Vogel and Risken [VI-13] focused on the Boson-Fock space, and proposed an estimation method of the density operator on this space based on an available quantum measurement that is referred to as Homodyne measurement. Any state on the Boson-Fock space is known to be described by a Wigner function. However, we cannot directly obtain this function through a quantum measurement. They proposed a method of estimating the Wigner function from the probability distribution \( p(x; \phi) \) of the data \( x \) of the measurement of the observable \( X_\phi \equiv Q \cos \phi + P \sin \phi \). That is, their method requires the probability densities \( p(x; \phi) \) of all angles \( \phi \). For this method, they derived a formula calculating the Wigner function from the probability distribution \( p(x; \phi) \) of the data \( x \) of the measurement

†Hayashi, Chapter 6 [II-6].
‡F. Tanaka, Investigation on Fisher metric on the classical statistical models and the quantum ones, master thesis, the University of Tokyo, 2004 (in Japanese).
of the observable $X_{\phi} \equiv Q \cos \phi + P \sin \phi$. Based on their results, Smithey et al. [VI-14] implemented the required measurements in a real optical system, and experimentally estimated Wigner function.

These researches estimated the true state by directly using the relation between the Wigner function and $p(x; \phi)$. However, there is a gap between the true distribution $p(x; \phi)$ and the experimentally estimated distribution $\hat{p}(x; \phi)$. Therefore, we need to consider how large the difference between the true state and the estimated state is. For this analysis, D’Ariano [Chap.33] constructed a function $K : p \mapsto K[p]$ transforming $p(x; \phi)$ to the density operator $\rho$. He focused on matrix elements $\langle n|K[p]|m \rangle$ based on the number state $|n\rangle$, and proved that the norm of the linear functional of $p \mapsto \langle n|K[p]|m \rangle$ is finite. This fact guarantees that the difference between the true matrix element and the estimated one is small if the difference between $p(x; \phi)$ and $\hat{p}(x; \phi)$ is also small. Furthermore, he discussed the case of finite quantum efficiency.

However, there is no reason to use only the linear function $K[\hat{p}]$ for the estimation of the density operator from the data $\hat{p}(x; \phi)$. If we use this method in the case when the true state is pure or nearly pure, the estimated density operator may have negative eigenvalues, i.e., may not be a density operator. Banaszek et al. [VI-15] proposed a method estimating the density operator $\hat{\rho}(X^n)$ from the finite data $X^n \equiv (X_1, \Phi_1), \ldots, (X_n, \Phi_n)$. Their method is regarded as a modification from the maximum likelihood estimation (MLE)$^\text{\S}$, which is a standard estimating method in classical statistical inference. Gill and Guță [VI-16] called this method the sieved maximum likelihood estimation, and proved that the norm $\|\rho - \hat{\rho}(X^n)\|_1$ between the true density operator $\rho$ and the estimated one $\hat{\rho}(X^n)$ converges to 0 with the probability 1 i.e., this convergence is almost sure. Furthermore, Artiles et al. [VI-17] discussed the implementation of this estimation, in more details.

The “tomographic” method can be also applied to the finite-dimensional case. In the above method, we obtain the probability distribution corresponding to the projection $E_\phi(x)$ at first, and obtain $\rho$ from a linear calculation, where $\int xE_\phi(x) \, dx$ is the spectral decomposition of $X_{\phi}$. Thus, this method is generalized to a more general setting on a finite-dimensional space $\mathcal{H}$ as follows: Let $X_1, \ldots, X_k$ be a basis of the linear space of matrix

$^\text{\S}$The definition of MLE is given in Sec. 6.2 of Chap. 7.
on $\mathcal{H}$, and $Y_1, \ldots, Y_k$ be the dual basis, i.e., satisfy
\[ \text{Tr} Y_i X_j = \delta_{i,j}. \]

Based on the above preparation, we measure observables $X_i$ several times, and obtain the respective average values $\mu_i$ and the estimated density matrix by [VI-19, eq.(19) of Chap. 34]
\[ \hat{\rho} = \sum_i \mu_i Y_i, \quad \text{Tr} \hat{\rho} X_i = \mu_i. \]

This method is called linear tomography.

However, if we use this method in the nearly pure state case, $\hat{\rho}$ may have negative eigenvalues. Another problem is that this method uses only the averages $\mu_i$ if the respective measured observables $X_i$. Usually, the empirical distribution is sufficient information to estimate the true distribution in statistics. If we use only the averages, we ignore much useful information and the accuracy of the estimation goes down. Thus, for a more precise estimation, we need to focus on the probability distribution under the measurements of the observables $X_1, \ldots, X_k$. Hradil [VI-20] and James et al. [Chap. 34] applied the maximum likelihood estimation to this problem. This method is called maximum likelihood tomography. James et al. [Chap. 34] compared the maximum likelihood tomography and linear tomography with a physical realizable setting, and demonstrated the superiority of the former. Their comparison is based on the application to the real optical system.

Moreover, Usami et al. [VI-18] applied maximum likelihood tomography to their experimental data, and checked its accuracy. They discussed the merit of the use of AIC (Akaike Information Criterion). As another method, Bužek et al. [VI-21] applied Bayes inference to the above setting under a prior distribution. This method is called Bayesian reconstruction, and they compared it with maximum entropy method.

**Further Reading**


¶Of course, this superiority is promised by the results of classical statistics.


CHAPTER 31

Optimal Quantum Clocks

Vladimír Bužek, R. Derka, and Serge Massar

Abstract. A quantum clock must satisfy two basic constraints. The first is a bound on the time resolution of the clock given by the difference between its maximum and minimum energy eigenvalues. The second follows from Holevo’s bound on how much classical information can be encoded in a quantum system. We show that asymptotically, as the dimension of the Hilbert space of the clock tends to infinity, both constraints can be satisfied simultaneously. The experimental realization of such an optimal quantum clock using trapped ions is discussed.

Recent technical advances in the laser cooling and trapping of ions suggest that coherent manipulations of trapped ions will be performed in the not too far future [13]. Apart from various important applications such as quantum information processing or improving high-precision spectroscopy these techniques also allow us to test fundamental concepts of quantum theory. In particular, much deeper insight into the problem of quantum measurement can be obtained.

In this Letter we study the problem of building an optimal quantum clock from an ensemble of \( N \) ions. To be specific let us assume an ion trap with \( N \) two-level ions all in the ground state \( |\Psi\rangle = |0\rangle \otimes \cdots \otimes |0\rangle \). This state is an eigenstate of the free Hamiltonian and thus cannot record time. Therefore the first step in building a clock is to bring the system to an initial state \( \hat{\Omega} \) which is not an energy eigenstate. For instance one can apply a Ramsey pulse whose shape and duration is chosen such that it puts all the ions in the product state

\[
\hat{\Omega}_{\text{prod}} = \hat{\rho} \otimes \cdots \otimes \hat{\rho} \quad ; \quad \hat{\rho} = \frac{1}{2} (|0\rangle + |1\rangle)(\langle 0| + \langle 1|).
\]

We shall also consider more general states, but shall always take them

to belong to the symmetric subspace of the $N$ ions. The basis vectors of this space will be denoted $|m\rangle$, $m = 0, 1, \ldots, N$. They are the completely symmetrized states of $N$ two-level ions with $m$ ions in the excited state and $(N-m)$ ions in the ground state. The states $|m\rangle$ have energy $E_m = m$ (this defines our unit of energy, setting $\hbar = 1$ then defines our unit of time).

The reason we can restrict ourselves to the symmetric subspace is that we can map any clock state onto the symmetric subspace without affecting its dynamics. Indeed consider an initial state $\Omega = |\psi\rangle \langle \psi|$ that does not belong to the symmetric subspace of the atoms. We can decompose $|\psi\rangle = \sum_m \sum_\alpha c_{m\alpha} |m, \alpha\rangle$ where $|m, \alpha\rangle$, $\alpha = 1, \ldots, \binom{N}{m}$ denote a basis of the states with energy $m$. Consider the unitary operator $\hat{U}$ that maps the state $|\psi\rangle$ onto the symmetric subspace without changing its energy: $\hat{U} \sum_\alpha c_{m\alpha} |m, \alpha\rangle = c_m |m\rangle$ where $|m\rangle$ is as before the symmetric state with energy $m$. Since $\hat{U}$ commutes with the Hamiltonian the performance of the clock based on $|\psi\rangle$ is identical to the clock whose initial state is the symmetric state $|\psi_{sym}\rangle = \hat{U} |\psi\rangle$.

After the preparation stage, the ions evolve in time according to the Hamiltonian evolution $\hat{\Omega}(t) = \hat{U}(t)\hat{\Omega}\hat{U}^\dagger(t)$, $\hat{U}(t) = \exp\{-it\hat{H}\}$. The task is to determine the elapsed time $t$ by carrying out a measurement on the ions. Note that because of the indeterminism of quantum mechanics it is impossible, given a single set of $N$ two-level ions, to determine the elapsed time with certainty. The best we can do is to estimate the elapsed time based on the result of a measurement on the system [4].

Making a good quantum clock requires a double optimization. First of all one can optimize the measurement. This aspect has been studied in detail in [4] where the best measuring strategy was derived. But one can also optimize the initial state $\hat{\Omega}$ of the system. It is this second optimization that is studied in this Letter.

Before turning to the problem of optimizing the initial state $\hat{\Omega}$, it is instructive to review the fundamental limitations on the performance of quantum clocks. Let us first consider a simple classical clock that can then be generalized to the quantum case. Our classical clock consists of a set of $n$ registers. Each register is either in the 0 or the 1 state. Thus the classical clock consists of $n$ bits, and can be in $2^n$ different states. The dynamics of the clock is as follows: the first register flips from 0 to 1 or from 1 to 0 every $2\pi 2^{-n}$, the second register flips every $2\pi 2^{-n+1}$, etc. The last register flips every $\pi$. This clock thus measures time modulo $2\pi$. Note that this clock has an inherent uncertainty since it cannot measure time with a precision better than $2\pi 2^{-n}$. Throughout this Letter the time uncertainty is defined
Optimal Quantum Clocks

as

\[ \Delta t^2 = [(t_{\text{estimate}} - t_{\text{true}}) \mod 2\pi]^2. \]  

(2)

For the classical clock \( \Delta t_{\text{class}} = \pi 2^{-n}/\sqrt{3} \).

It is straightforward to replace the classical clock by a quantum version. The quantum clock consists of \( n \) two-level systems (qubits). The first qubit has an energy splitting between the levels of \( 2^{n-1} \) so that it has the same period as the corresponding classical register. The second qubit has an energy splitting of \( 2^{n-2} \) and so on up to the last qubit that has an energy splitting of 1. Considered together these \( n \) qubits constitute a quantum system with \( 2^{n} \) equally spaced energy levels.

The mapping between the Hilbert space of this abstract quantum clock and the symmetric subspace of \( N \) two-level ions is straightforward when \( N + 1 = 2^n \). Indeed in this case the dimension and energy spectrum of both Hilbert spaces coincide. Note, however, that this is not a mapping between the qubits of the clock and the ions individually, but between energy eigenstates. This comparison between classical and quantum clocks suggests that a quantum clock built out of \( N \) ions cannot behave better than a classical clock built out of \( \ln(N+1) \) registers. That this is indeed the case follows from two fundamental constraints:

The first constraint is a bound on the time resolution of the clock that results from its energy spectrum. Indeed the time-energy uncertainty \([7, 2]\)

\[ \Delta t \Delta E \geq \frac{1}{2}, \]  

(3)

where \( \Delta E^2 = \text{Tr}(\hat{H}^2 \hat{\Omega}) - [\text{Tr}(\hat{H} \hat{\Omega})]^2 \) relates the uncertainty in the estimated time [defined by Eq. (2)] to the spread in energy of the clock. In the present case there is a state with a maximum energy uncertainty, namely \( |\psi_+\rangle = (1/\sqrt{2})(|N\rangle + |0\rangle) \) for which \( \Delta E = N/2 \). Inserting in Eq. (3) shows that for any clock built out of \( N \) ions

\[ \Delta t \geq \frac{1}{N}. \]  

(4)

Note that one cannot attain equality in Eq. (3). Indeed the state \( |\psi_+\rangle \) evolves with a period \( 2\pi/N \), hence it necessarily has a large time uncertainty. Thus Eq. (4) is only a lower bound and maximizing the energy uncertainty as in [1] is not necessarily an optimal procedure.
The second fundamental constraint the clock must obey is a bound on the total information it can carry. Indeed Holevo [5] has shown that one cannot encode and subsequently retrieve reliably more than \( n \) bits of classical information into \( n \) qubits. Letting a clock evolve for a given time interval \( t \) can be viewed as trying to encode information about the classical variable \( t \) into the state of the clock. Hence a measurement on our model clock [consisting of \( n = \ln(N+1) \) qubits] cannot retrieve more than \( \ln(N+1) \) bits of information about \( t \).

Together these two bounds imply that the quantum clock cannot perform better than the corresponding classical clock: it cannot carry more information and it cannot have better resolution. But is it possible, by making an optimal measurement and choosing in an astute manner the initial state of the clock, to make a quantum clock with similar performances to the classical one? Our main result is to show that this is indeed the case for clocks built out of symmetric states of \( N \) ions and to provide an algorithm for constructing such an optimal clock.

The problem of constructing quantum clocks has been considered previously in [10, 9]. However, the best clocks considered in these papers are based on the phase state \( |\Psi_0\rangle \) described below. As we shall see for these clocks the time uncertainty is very large \( \Delta t \simeq (1/\sqrt{N}) \) and is very far from reaching equality in Eq. (4). Recently, Vaidman and Belkind [12] considered the problem of a clock for which equality holds in Eq. (3). They showed that in the limit of large \( N \) the product states satisfy this condition. However, for the product state the energy uncertainty is very small: \( \Delta E = \sqrt{N}/2 \) hence they also do not saturate Eq. (4). Furthermore, clocks based on product state also do not attain Holevo’s bound.

A similar approach to the one used here, namely optimizing both the initial state and the measurement on a system of \( N \) ions was considered in [6] with the aim of using the ions as an improved frequency standard. This problem can be rephrased in the following way: one disposes of a classical but noisy clock which provides some \textit{a priori} knowledge about the time \( t \) and one wants to improve the knowledge of \( t \) by using the \( N \) ions. On the other hand, in the present Letter we suppose that there is no prior knowledge about \( t \). The other difference with the present work is that our aim is to study the fundamental structure of quantum mechanics. We therefore neglect the effect of noise during the preparation and measurement stages and decoherence during the evolution. On the other hand, taking these effects into account was central to [6].
In order to find how to build an optimal clock we must delve in detail into the functioning of this device. We first recall Holevo’s results concerning the optimal measurement strategy [4]. The measurement is described by a positive operator measurement (POVM), that is a set \( \{\hat{O}_r\}_{r=1}^R \) of positive Hermitian operators such that \( \sum_r \hat{O}_r = \hat{1} \). To each outcome \( r \) of the measurement we associate an estimate \( t_r \) of the time elapsed. The difference between the estimated time \( t_r \) and the true time \( t \) is measured by a cost function \( f(t_r - t) \). Here we note that because of the periodicity of the clock, \( f \) has to be periodic. We also take \( f(t) \) to be an even function. The task is to minimize the mean value of the cost function

\[
\bar{f} = \sum_r \int_0^{2\pi} \text{Tr}[\hat{O}_r \hat{\Omega}(t)] f(t_r - t) \frac{dt}{2\pi}.
\]

To proceed we expand the cost function in Fourier series:

\[
f(t) = w_0 - \sum_{k=1}^{\infty} w_k \cos kt.
\]

The essential hypothesis made by Holevo is positivity of the Fourier coefficients: \( w_k \geq 0 \), \( k = 1, 2, ... \). He also supposes that the initial state \( \hat{\Omega} = |\omega\rangle\langle\omega|, |\omega\rangle = \sum_m a_m |m\rangle \) is a pure state (and we make a phase convention such that \( a_m \) is real and positive). He then shows that

\[
\bar{f} \geq w_0 - \frac{1}{2} \sum_{k=1}^{\infty} w_k \sum_{|m-m'|=k} a_m a_{m'},
\]

and equality is attained only if the measurement is of the form

\[
\hat{O}_r = p_r |\Psi_r\rangle\langle\Psi_r| \quad ; \quad p_r \geq 0 \quad ; \quad |\Psi_r\rangle = e^{it_r\hat{H}} |\Psi_0\rangle,
\]

\[
|\Psi_0\rangle = \frac{1}{\sqrt{N+1}} \sum_{m=0}^N |m\rangle,
\]

with the completeness relation \( \sum_r \hat{O}_r = \hat{1} \).

Several remarks about this result are called for:

1 Holevo supposed that the initial state is a pure state. If the initial state is mixed, \( \hat{\Omega} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \), then one finds that the corresponding cost is bounded by the average of the bounds Eq. (7). This shows that in building a good quantum clock one should always take the initial state to be pure.
2 Holevo considered covariant measurements in which the times \( t_r \) takes the continuum of values between 0 and \( 2\pi \). But as shown in [3] the completeness relation can also be satisfied by taking a discreet set of times \( t_j = \frac{2\pi j}{N+1} \), \( j = 0, \ldots, N \). These “phase” states \(|\Psi_j\rangle\) [8] form an orthonormal basis of the Hilbert space, and this measurement is therefore a von Neumann measurement. Thus it is not necessary to use an ancilla to make the optimal measurement in this case.

3 In Eq. (7) only the values \( k = 0, \ldots, N \) intervene because of the condition \(|m - m'| = k\). That is, only the first \( N + 1 \) Fourier coefficients of the cost function are meaningful.

4 Because of the condition of positivity of \( w_k \), not all cost functions are covered by this result, but several important examples are \( 4\sin^2 \frac{t}{2} = 2(1 - \cos t) \), \(|\sin \frac{2t}{N}\rangle \), \(-\delta[t(\mod 2\pi)]\). The most notable absence from this list is the quadratic deviation \( t^2 \) [as defined in Eq. (2)] but it can be well approximated by the the first cost function since \( 4\sin^2 \frac{t}{2} \approx t^2 \) for \(|t| \ll \pi\).

We now turn back to the central problem of this Letter, namely how to optimize the initial state of the system so that the time estimate is as good as possible. This corresponds to minimizing the right-hand side of Eq. (7) with respect to \( a_m \). To this end let us define the matrix

\[
F_{mm'} = w_0 \delta_{m,m'} - \frac{1}{2} \sum_{k=1}^{N} w_k (\delta_{m,m'+k} + \delta_{m+k,m'}). 
\]

We must then minimize \( \bar{f} = a^T \hat{F} a \) under the condition \( a^T a = 1 \). Using a Lagrange multiplier we find the eigenvalue equation \((\hat{F} - \lambda \hat{1})a = 0\), and the task is therefore to find the smallest eigenvalue and eigenvector of \( \hat{F} \).

Let us first consider the cost function \( f = 4\sin^2 \frac{t}{2} \). The advantage of this cost function is that the matrix \( F \) is particularly simple in this case. Furthermore, for errors much smaller then \( 2\pi \), \( f \) and \( \Delta t^2 \) as defined in Eq. (2) coincide. Hence the first constraint on clock resolution Eq. (4) can be approximately replaced by \( \bar{f} \geq 1/N^2 \).

If the initial state is the product state Eq. (1) then the mean cost is given by the expression \( \bar{f} = 2[1 - 2^{-N} \sum_{i=0}^{N-1} \sqrt{\binom{N}{i}}(\frac{N-i}{N+1})] \) which for large \( N \) decreases as \( \bar{f} \approx 1/N \). If the initial state is the phase state \(|\Psi_0\rangle\), then direct calculation shows that \( \bar{f} = \frac{2}{N+1} \). Thus for both states \( \Delta t \approx N^{-1/2} \) and one is very far from attaining equality in Eq. (4).
However, neither of these two states is optimal. To find the optimum we note that the matrix \( \hat{F} = 2 \delta_{mm'} - \delta_{mm'+1} - \delta_{m+1m'} \) can be viewed as the discretized second derivative operator \( \hat{F} \simeq -d^2/dx^2 \) with von Neumann boundary conditions. The lowest eigenvalue of \( \hat{F} \) is therefore approximately \( \lambda_{\text{min}} \simeq \pi^2/(N + 1)^2 \) and the corresponding eigenvector is

\[
|\Psi_{\text{opt}}\rangle \simeq \frac{\sqrt{2}}{\sqrt{N + 1}} \sum_m \sin \frac{\pi (m + 1/2)}{N + 1} |m\rangle. \tag{10}
\]

Thus in this case the cost decreases for large \( N \) as \( \tilde{f}_{\text{opt}} \simeq \frac{\pi^2}{(N+1)^2} \) corresponding to \( \Delta t_{\text{opt}} \simeq \frac{\pi}{(N+1)} \). Therefore, up to a factor of \( \pi \) the optimal clock attains the bound Eq. (4).

It is also interesting to consider the situation where the cost function is the delta function \( f = -\delta[t (\mod 2\pi)] \). In this case \( F_{mm'} = -\frac{1}{2\pi} \) for all \( m, m' \). One checks that the phase state \( |\Psi_0\rangle \) is the eigenvector of \( \hat{F} \) with minimal eigenvalue \( \lambda = -\frac{N}{2\pi} \). Note that one could also have taken a smeared delta function since only the first \( N + 1 \) terms intervene in Eq. (7). The smeared delta function is approximately zero everywhere except in an interval of about \( 1/N \) around zero where it is equal to \( N \). Thus for this cost function one wants to maximize the frequency with which the estimated value of \( t \) is within about \( \frac{1}{N} \) of the true value. But there is no extra cost if the estimated value is very far from the true one. It is for this reason that taking \( |\Psi_0\rangle \) as the initial state when the cost function is \( 4 \sin^2 t/2 \) is bad since making estimates that are wildly off is strongly penalized in that case.

The mean value of a cost function gives only very partial information about the sensitivity of a clock. The full information is encoded in the probability \( P(t|t_r) = P(\hat{O}_r|t)P(t)/P(\hat{O}_r) = P(\hat{O}_r|t)\frac{N+1}{2\pi} \) that the true time is \( t \) given that the readout of the measurement is \( t_r \). In the case of the optimal state \( |\Psi_{\text{opt}}\rangle \), one finds that

\[
P_{\text{opt}}(t|t_r) \simeq \mathcal{N} \frac{[1 + \cos(N + 1)T][1 + \cos T]}{[1 - \cos(T + \frac{\pi}{N+1})][1 - \cos(T - \frac{\pi}{N+1})]}, \tag{11}
\]

where \( \mathcal{N} \simeq \frac{\pi}{2(N+1)^2} \) and \( T = t - t_r \). This distribution (see Fig. 1) has a central peak of width \( 3\pi/(N + 1) \) and tails which decrease for \( 2\pi g|t - t_r|gN^{-1} \) as \( P_{\text{opt}}(t|t_r) \simeq N^{-3}|t - t_r|^{-4} \).

For the phase state \( |\Psi_0\rangle \) one finds

\[
P_{\Psi_0}(t|t_r) = \frac{1}{2\pi(N + 1)} \frac{[1 - \cos(N + 1)T]}{(1 - \cos T)}. \tag{12}
\]
Fig. 1. A plot of the a posteriori probability $P(t|t_r)$ that the time was $t$ given that the measurement yielded outcome $t_r = \pi$ for different initial states ($N = 20$). The dotted line corresponds to the product state Eq. (1), the solid line to the phase state $|\Psi_0\rangle$ (8), and the dashed line to the optimal state $|\Psi_{\text{opt}}\rangle$ (10).

This distribution has a slightly tighter central peak of width $2\pi/(N + 1)$ but the tails of the distribution decrease as $N^{-1}|t - t_r|^{-2}$. It is these slowly decreasing tails that give the main contribution to $\Delta t \simeq N^{-1/2}$.

For the product state Eq. (1) the distribution $P_{\text{prod}}(t|t_r)$ has a very wide central peak of width $\simeq \frac{1}{\sqrt{N}}$. In this case it is the wide central peak that gives rise to the large time uncertainty.

Using these distributions it is possible to calculate the number of bits of information about time that is encoded in the outcomes of the measurement. This is given by the mutual information

$$I = - \int dt P(t) \ln P(t) + \sum_r P(t_r) \int dt P(t|t_r) \ln P(t|t_r).$$

In all cases the integral in the second term is dominated by the central peak. Thus one finds that for the product state only $\frac{1}{2} \ln N$ bits of information about time are obtained, whereas for both the states $|\Psi_0\rangle$ and $|\Psi_{\text{opt}}\rangle$ one obtains $\ln N$ bits, thereby saturating Holevo’s bound.

In summary we have seen that for different cost functions there are different optimal clocks. There is, of course, no cost function that is in an absolute sense better than another, and the choice of a particular cost
function depends on the physical context. Nevertheless, the experimental realization of a quantum clock based on the state \( |\Psi_{opt}\rangle \) seems particularly desirable because it combines the attractive features that the \emph{a posteriori} probability \( P(t|t_r) \) has a tight central peak and rapidly decreasing tails.

Carrying out such an experiment with trapped ions presents two main difficulties. The first is the preparation of the initial state \( \hat{\Omega} \). Such coherent manipulation of trapped ions is one of the current experimental challenges. A possibly important simplifying feature of the quantum clock is that since it is symmetric in the \( N \) ions one does not need to address each ion individually. The second problem is to realize the optimal phase measurement as discussed in the present Letter. Recent experiments [11] suggest that this type of coherent preparation and specific projective measurements are possible for systems with a moderate number of trapped ions. Therefore, one may hope that it will be feasible to make optimal quantum clocks in the not too distant future.

**Acknowledgements**

Our special thanks to Lev Vaidman for his criticism and careful rereading of the manuscript. We also thank Win van Dam, Susana Huelga, Peter Knight and Christopher Monroe for helpful discussions. This work was in part supported by the Royal Society and by the Slovak Academy of Sciences. V. B. and S. M. thank the Benasque Center for Physics where part of this work was carried out. S.M. would like to thank Utrecht University where most of this work was carried out; he is a chercheur qualifié du FNRS.

**References**


CHAPTER 32

Quantum Channel Identification Problem

Akio Fujiwara

Abstract. This paper explores an application of quantum entanglement. The problem treated here is the quantum channel identification problem: given a parametric family \( \{ \Gamma_\theta \}_{\theta} \) of quantum channels, find the best strategy of estimating the true value of the parameter \( \theta \). As a simple example, we study the estimation problem of the isotropic depolarization parameter \( \theta \) for a two-level quantum system \( \mathcal{H} \simeq \mathbb{C}^2 \). In the framework of noncommutative statistics, it is shown that the optimal input state on \( \mathcal{H} \otimes \mathcal{H} \) to the channel exhibits a transition-like behavior according to the value of the parameter \( \theta \).

Let \( \mathcal{H} \) be a Hilbert space that represents the physical system of interest and let \( S(\mathcal{H}) \) be the set of density operators on \( \mathcal{H} \). It is well known [4] that a dynamical change \( \Gamma : S(\mathcal{H}) \to S(\mathcal{H}) \) of the physical system, called a quantum channel, is represented by a trace-preserving completely positive map. But how can we identify the quantum channel that we have in a laboratory? A general scheme may be as follows: input a well-prepared state \( \sigma \) to the quantum channel and estimate the dynamical change \( \sigma \mapsto \Gamma(\sigma) \) by performing a certain measurement on the output state \( \Gamma(\sigma) \). It is then natural to inquire what is the best strategy of estimating a quantum channel. The purpose of this paper is to study this problem from a noncommutative statistical point of view. For mathematical simplicity, we restrict ourselves to the case in which the quantum channel to be identified lies in a smooth parametric family \( \{ \Gamma_\theta; \theta = (\theta_1, ..., \theta_n) \in \Theta \} \) of quantum channels. When \( \mathcal{H} \) is finite-dimensional, this is not an essential restriction [1].

Once an input state \( \sigma \) for the channel is fixed, we have a parametric family \( \{ \Gamma_\theta(\sigma) \}_{\theta \in \Theta} \) of output states, and as long as the parametrization \( \theta \mapsto \Gamma_\theta(\sigma) \) is nondegenerate, the problem of estimating the quantum channel is reduced to a parameter estimation problem for the noncommutative statistical model \( \{ \Gamma_\theta(\sigma) \}_{\theta \in \Theta} \). As a consequence, the parameter estimation problem for a family \( \{ \Gamma_\theta \}_{\theta \in \Theta} \) of quantum channels amounts to finding an

optimal input state $\sigma$ for the channel and an optimal estimator for the parametric family $\{\Gamma_\theta(\sigma)\}_\theta$ of output states. One may imagine that this problem does not exceed the realm of conventional quantum estimation theory [2, 3]. But, in fact, it opens a new field of research in noncommutative statistics.

Since each channel $\Gamma_\theta$ is completely positive, it can be extended to the composite quantum system $\mathcal{H} \otimes \mathcal{H}$. In view of statistical parameter estimation, there are two essentially different extensions that have the same parametrization $\theta$ as $\Gamma_\theta$: one is $\Gamma_\theta \otimes I : S(\mathcal{H} \otimes \mathcal{H}) \to S(\mathcal{H} \otimes \mathcal{H})$, where $I$ denotes the identity channel, and the other is $\Gamma_\theta \otimes \Gamma_\theta : S(\mathcal{H} \otimes \mathcal{H}) \to S(\mathcal{H} \otimes \mathcal{H})$. A question arises naturally: what happens when we use an entangled state as an input to the extended channel? In what follows, we demonstrate a somewhat nontrivial aspect of this problem.

Let $\mathcal{H} := \mathbb{C}^2$ and let the channel $\Gamma_\theta : S(\mathcal{H}) \to S(\mathcal{H})$ be defined by

$$\Gamma_\theta \left( \frac{1}{2} \begin{bmatrix} 1 + z & x - iy \\ x + iy & 1 - z \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} 1 + \theta z & \theta(x - iy) \\ \theta(x + iy) & 1 - \theta z \end{bmatrix}.$$  

The parameter $\theta$ represents the magnitude of isotropic depolarization. The channel can be uniquely extended on the $2 \times 2$ matrix algebra $\mathbb{C}^{2 \times 2}$ as follows.

$$\Gamma_\theta \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 0 & 2\theta \\ 0 & 0 \end{bmatrix}, \quad \Gamma_\theta \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 + \theta & 0 \\ 0 & 1 - \theta \end{bmatrix},$$

$$\Gamma_\theta \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 0 & 0 \\ 2\theta & 0 \end{bmatrix}, \quad \Gamma_\theta \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 - \theta & 0 \\ 0 & 1 + \theta \end{bmatrix}.$$  

To ensure that $\Gamma_\theta$ is completely positive, the parameter $\theta$ must lie in the closed interval $\Theta := [-\frac{1}{3}, 1]$. (See [1].) We thus have a one-parameter family $\{\Gamma_\theta; \theta \in \Theta\}$ of quantum channels, and our task is to estimate the true value of $\theta$.

Before proceeding to the parameter estimation for $\{\Gamma_\theta\}_\theta$, we give a brief account of the one-parameter quantum estimation theory for density operators. (Consult [2] or [3] for details.) Given a one-parameter family $\{\rho_\theta\}_\theta$ of density operators, an estimator for the parameter $\theta$ is represented by a Hermitian operator $T$, normally with a requirement that the estimator should be (locally) unbiased: that is, if the system is in the state $\rho_\theta$, then the expectation $E_\theta[T] := \text{Tr}\rho_\theta T$ of the estimator $T$ should be identical to $\theta$. It is easy to show that every (locally) unbiased estimator $T$ for the parameter $\theta$ satisfies the quantum Cramér-Rao inequality $V_\theta[T] \geq (J_\theta)^{-1}$, where $V_\theta[T] := \text{Tr}\rho_\theta (T - \theta)^2$ is the variance of estimator $T$. 


and \( J_\theta := J(\rho_\theta) := \text{Tr}\rho_\theta(L_\theta)^2 \) is the quantum Fisher information with \( L_\theta \) the symmetric logarithmic derivative (SLD), i.e., the Hermitian operator that satisfies the equation

\[
\frac{d\rho_\theta}{d\theta} = \frac{1}{2}(L_\theta\rho_\theta + \rho_\theta L_\theta).
\]

It is important to notice that the lower bound \((J_\theta)^{-1}\) in the quantum Cramér-Rao inequality is achievable (at least locally). In other words, the inverse of the SLD Fisher information gives the ultimate limit of estimation. As a consequence, the larger the SLD Fisher information is, the more accurately we can estimate the parameter \( \theta \).

Let us return to the parameter estimation problem for the one-parameter family \( \{\Gamma_\theta\}_\theta \) of quantum channels. Taking account of the above-mentioned one-parameter estimation theory for density operators, our task is reduced to finding an optimal input for the channel that maximizes the SLD Fisher information of the corresponding parametric family of output states.

We start with the maximization of the SLD Fisher information of the family \( \{\Gamma_\theta(\sigma)\}_\theta \) with respect to the input state \( \sigma \in \mathcal{S}(\mathcal{H}) \). An important observation is that the maximum is attained by a pure state. To see this, it suffices to prove the convexity of the SLD Fisher information, i.e., if \( \rho_\theta = \lambda\sigma_\theta + (1 - \lambda)\tau_\theta \) for a constant \( \lambda \) between 0 and 1, then \( J(\rho_\theta) \leq \lambda J(\sigma_\theta) + (1 - \lambda)J(\tau_\theta) \). Let us introduce the states \( \tilde{\rho}_\theta := \lambda\sigma_\theta \oplus (1 - \lambda)\tau_\theta \) on \( \mathcal{H} \oplus \mathcal{H} \). It is easy to show that \( J(\tilde{\rho}_\theta) = \lambda J(\sigma_\theta) + (1 - \lambda)J(\tau_\theta) \). By identifying \( \mathcal{H} \oplus \mathcal{H} \) with \( \mathbb{C}^2 \otimes \mathcal{H} \), we consider the partial trace \( \text{Tr}_{\mathbb{C}^2} : \mathcal{S}(\mathcal{H} \oplus \mathcal{H}) \to \mathcal{S}(\mathcal{H}) \). Since \( \text{Tr}_{\mathbb{C}^2} \) is a stochastic (i.e., trace preserving completely positive) map, the monotonicity of the SLD Fisher information with respect to a stochastic map [5] shows that \( J(\tilde{\rho}_\theta) \geq J(\text{Tr}_{\mathbb{C}^2} \tilde{\rho}_\theta) = J(\rho_\theta) \). This completes the proof of the convexity of the SLD Fisher information. Now we are ready to specify an optimal input state. Since our channel \( \Gamma_\theta \) is unitarily invariant (i.e., isotropic in the Stokes parameter space), we can take without loss of generality the optimal input to be \( \sigma = |e\rangle\langle e| \) where \( \langle e| = (1, 0) \). The corresponding output state \( \rho_\theta := \Gamma_\theta(\sigma) \) is

\[
\rho_\theta = \frac{1}{2} \begin{bmatrix} 1 + \theta & 0 \\ 0 & 1 - \theta \end{bmatrix}.
\]

Since the state \( \rho_\theta \) is isomorphic to the classical coin flipping in which “heads” occur with probability \((1 + \theta)/2\), the SLD Fisher information be-
comes

\[ J_\theta = \frac{1}{1 - \theta^2}. \]

We next study the extended channel \( \Gamma_\theta \otimes I : \mathcal{S}(\mathcal{H} \otimes \mathcal{H}) \to \mathcal{S}(\mathcal{H} \otimes \mathcal{H}) \). In this case we can use a possibly entangled state as the input. For the same reason as above, we can take the input to be a pure state: \( \hat{\sigma} = |\psi\rangle\langle\psi| \) where \( \psi \in \mathcal{H} \otimes \mathcal{H} \). By the Schmidt decomposition, the vector \( \psi \) is represented as

\[ |\psi\rangle = \sqrt{x} |e_1\rangle|f_1\rangle + \sqrt{1-x} |e_2\rangle|f_2\rangle, \quad (1) \]

where \( x \) is a real number between 0 and 1, and \( \{e_1, e_2\} \) and \( \{f_1, f_2\} \) are orthonormal bases of \( \mathcal{H} = \mathbb{C}^2 \). Since the channels \( \Gamma_\theta \) and \( I \) are both unitarily invariant, we can assume without loss of generality that the optimal input takes the form (1) with \( \langle e_1 | = \langle f_1 | = (0, 1) \) and \( \langle e_2 | = \langle f_2 | = (1, 0) \).

The constant \( x \) remains to be determined. The corresponding output state \( \hat{\rho}_\theta := \Gamma_\theta \otimes I(\hat{\sigma}) \) becomes

\[
\hat{\rho}_\theta = \frac{1}{2} \begin{bmatrix}
(1 - x)(1 + \theta) & 0 & 0 & 2\sqrt{x(1-x)} \theta \\
0 & x(1-\theta) & 0 & 0 \\
0 & 0 & (1-x)(1-\theta) & 0 \\
2\sqrt{x(1-x)} \theta & 0 & 0 & x(1+\theta)
\end{bmatrix}.
\]

The SLD for the family \( \{\hat{\rho}_\theta\}_\theta \) is given by

\[
\hat{L}_\theta = \begin{bmatrix}
\frac{1 + 2\theta - 3\theta^2 - 8\theta x}{(1-\theta^2)(1+3\theta)} & 0 & 0 & \frac{4\sqrt{x(1-x)}}{(1-\theta)(1+3\theta)} \\
0 & \frac{1}{1-\theta} & 0 & 0 \\
0 & 0 & -\frac{1}{1-\theta} & 0 \\
\frac{4\sqrt{x(1-x)}}{(1-\theta)(1+3\theta)} & 0 & 0 & \frac{1 - 6\theta - 3\theta^2 + 8\theta x}{(1-\theta^2)(1+3\theta)}
\end{bmatrix},
\]

and the SLD Fisher information is

\[
\hat{J}_\theta = \frac{1 + 3\theta + 8x(1-x)}{(1-\theta^2)(1+3\theta)}.
\]

When \( x = 0 \) or \( 1 \), the above SLD Fisher information \( \hat{J}_\theta \) is identical to \( J_\theta \). This is a matter of course: the input state is disentangled in this case and no information about the parameter \( \theta \) is available via the independent channel \( I \). When \( x \neq 0 \) and \( \neq 1 \), the SLD Fisher information \( \hat{J}_\theta \) diverges at \( \theta = 1 \) and \(-\frac{1}{3}\). This is because the complete positivity of the channel \( \Gamma_\theta \) breaks across these values. Now let us specify the optimal input state. For every
\[ \begin{aligned}
\theta, \text{ the SLD Fisher information } J_\theta \text{ takes the maximum } 3/(1 - \theta)(1 + 3\theta) \text{ at } x = \frac{1}{2}. \text{ Therefore the optimal input for the channel } \Gamma_\theta \otimes I \text{ is the maximally entangled state. The implication of this result is profound: although we use the channel } \Gamma_\theta \text{ only once, extra information about the channel is obtained via entanglement of the input state. In particular, the use of entanglement improves exceedingly the performance of estimation as } \theta \text{ approaches } -\frac{1}{3}.
\end{aligned} \]

Let us proceed to the analysis of the other extended channel \( \Gamma_\theta \otimes \Gamma_\theta : \mathcal{S}(\mathcal{H} \otimes \mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H} \otimes \mathcal{H}) \). As before, we can take the input to be a pure state \( \tilde{\sigma} = |\psi\rangle \langle \psi| \) where \( \psi \) is given by Eq. (1) with \( \langle e_1 \rangle = \langle f_1 \rangle = (0,1) \) and \( \langle e_2 \rangle = \langle f_2 \rangle = (1,0) \). The corresponding output state \( \tilde{\rho}_\theta := \Gamma_\theta \otimes \Gamma_\theta(\tilde{\sigma}) \) becomes

\[ \tilde{\rho}_\theta = \frac{1}{4} \begin{bmatrix}
    x(1-\theta)^2 + (1-x)(1+\theta)^2 & 0 & 0 & 4\sqrt{x(1-x)} \theta^2 \\
    0 & 1-\theta^2 & 0 & 0 \\
    0 & 0 & 1-\theta^2 & 0 \\
    4\sqrt{x(1-x)} \theta^2 & 0 & 0 & x(1+\theta)^2 + (1-x)(1-\theta)^2
\end{bmatrix}. \]

Since the SLD for the family \( \{ \tilde{\rho}_\theta \}_\theta \) is too complicated to write down, we give the SLD Fisher information only.

\[ J_\theta = \frac{4\theta^4 + 5\theta^2 - 1}{2\theta^2(1 - \theta^4)} + \frac{8\theta^2 x(1-x)}{1-\theta^4} + \frac{1-\theta^2}{2\theta^2(1+\theta^2)(1-\theta^2 + 16\theta^2 x(1-x))}. \]

When \( x = 0 \) or \( 1 \), the above SLD Fisher information \( J_\theta \) becomes \( 2/(1 - \theta^2) \), which precisely doubles the \( J_\theta \). Again this is a matter of course: the input state is disentangled in this case and the same amount of information about the parameter \( \theta \) is obtained per independent use of the channel \( \Gamma_\theta \).

When \( x \neq 0 \) and \( \neq 1 \), the SLD Fisher information \( J_\theta \) diverges at \( \theta = 1 \) but does not at \( \theta = -\frac{1}{3} \). This is because the requirement of positivity for the channel \( \Gamma_\theta \otimes \Gamma_\theta \) is strictly weaker than that for the channel \( \Gamma_\theta \otimes I \) (i.e., the complete positivity for \( \Gamma_\theta \)). Now we examine a rather unexpected behavior of the optimal input state. For \( \frac{1}{\sqrt{3}} \leq \theta < 1 \), the SLD Fisher information \( J_\theta \) takes the maximum \( 12\theta^2/(1 - \theta^2)(1 + 3\theta^2) \) at \( x = \frac{1}{2} \), while for \( -\frac{1}{3} \leq \theta \leq \frac{1}{\sqrt{3}} \), it takes the maximum \( 2/(1 - \theta^2) \) at \( x = 0 \) and 1. (See Fig. 1.) Namely, the optimal input state “jumps” from the maximally entangled state to a disentangled state at \( \theta = 1/\sqrt{3} \). It is surprising that the seemingly homogeneous family \( \{ \Gamma_\theta \}_\theta \) of depolarization channels involves a transition-like behavior.

Finally we mention the possibility of extending the channel \( \Gamma_\theta \) in the form \( \Gamma_\theta \otimes \Gamma' \) where \( \Gamma' \) is a channel that is known to the observer and is independent of \( \theta \). Since the channel \( \Gamma_\theta \otimes \Gamma' \) is decomposed into \( (I \otimes \Gamma')(\Gamma_\theta \otimes \Gamma') \)
Fig. 1. SLD Fisher information $\tilde{J}_\theta$ versus $x$, for $\theta = 0.7$ (dashed), $\theta = 1/\sqrt{3}$ (solid), and $\theta = 0.3$ (chained).

$I$), the monotonicity argument for the SLD Fisher information with respect to a stochastic map allows us to deduce that the best choice of the channel $\Gamma'$ is the identity channel.

To conclude, among those we have considered on the second extension $\mathcal{H} \otimes \mathcal{H}$ of the quantum system, the best strategy of estimating the isotropic depolarization parameter $\theta$ is the following. For $\frac{1}{\sqrt{3}} \leq \theta \leq 1$, use $\Gamma_\theta \otimes \Gamma_\theta$ and input a maximally entangled state on $\mathcal{H} \otimes \mathcal{H}$; for $\frac{1}{3} \leq \theta \leq \frac{1}{\sqrt{3}}$, use $\Gamma_\theta$ twice independently and input any pure state on $\mathcal{H}$ each time; for $-\frac{1}{3} \leq \theta \leq \frac{1}{3}$, use $\Gamma_\theta \otimes I$ and input a maximally entangled state on $\mathcal{H} \otimes \mathcal{H}$.

We have demonstrated a nontrivial aspect of statistical estimation problem for a quantum channel. Other problems, such as the use of the $n$th extension $\mathcal{H} \otimes^n$ and its asymptotics, or the multi parameter quantum channel estimation, will be presented elsewhere.

References
[5] Intuitively speaking, the monotonicity of the SLD Fisher information implies that the distance of two nearby quantum states measured by the SLD Fisher metric shrinks as they are mapped by a stochastic map. For a complete
CHAPTER 33

Homodyning as Universal Detection

Giacomo Mauro D’Ariano

Abstract. Homodyne tomography — i.e., homodyning while scanning the local oscillator phase — is now a well assessed method for “measuring” the quantum state. In this paper I will show how it can be used as a kind of universal detection, for measuring generic field operators, however at expense of some additional noise. The general class of field operators that can be measured in this way is presented, and includes also operators that are inaccessible to heterodyne detection. The noise from tomographical homodyning is compared to that from heterodyning, for those operators that can be measured in both ways. It turns out that for some operators homodyning is better than heterodyning when the mean photon number is sufficiently small. Finally, the robustness of the method to additive phase-insensitive noise is analyzed. It is shown that just half photon of thermal noise would spoil the measurement completely.

1. Introduction

Homodyne tomography is the only viable method currently known for determining the detailed state of a quantum harmonic oscillator — a mode of the electromagnetic field. The state measurement is achieved by repeating many homodyne measurements at different phases $\phi$ with respect to the local oscillator (LO). The experimental work of the group in Eugene-Oregon [18] undoubtedly established the feasibility of the method, even though the earlier data analysis were based on a filtered procedure that affected the results with systematic errors. Later, the theoretical group in Pavia-Italy presented an exact reconstruction algorithm [7], which is the method currently adopted in actual experiments (see, for example, Refs. [16] and [17]). The reconstruction algorithm of Ref. [7] was later greatly simplified [5], so that it was possible also to recognize the feasibility of the method even for nonideal quantum efficiency $\eta < 1$ at the homodyne detector, and, at the

same time, establishing lower bounds for $\eta$ for any given matrix representation. After these first results, further theoretical progress has been made, understanding the mechanisms that underly the generation of statistical errors [2], thus limiting the sensitivity of the method. More recently, for $\eta = 1$ non trivial factorization formulas have been recognized [19, 14] for the “pattern functions” [15] that are necessary to reconstruct the photon statistics.

In this paper I will show how homodyne tomography can also be used as a method for measuring generic field operators. In fact, due to statistical errors, the measured matrix elements cannot be used to obtain expectations of field operators, and a different algorithm for analyzing homodyne data is needed suited to the particular field operator whose expectation one wants to estimate. Here, I will present an algorithm valid for any operator that admits a normal ordered expansion, giving the general class of operators that can be measured in this way, also as a function of the quantum efficiency $\eta$. Hence, from the same bunch of homodyne experimental data, now one can obtain not only the density matrix of the state, but also the expectation value of various field operators, including some operators that are inaccessible to heterodyne detection. However, the price to pay for such detection flexibility is that all measured quantities will be affected by noise. But, if one compares this noise with that from heterodyning (for those operators that can be measured in both ways), it turns out that for some operators homodyning is less noisy than heterodyning, at least for small mean photon numbers.

Finally, I will show that the method of homodyne tomography is quite robust to sources of additive noise. Focusing attention on the most common situation in which the noise is Gaussian and independent on the LO phase, I will show that this kind of noise produces the same effect of nonunit quantum efficiency at detectors. Generalizing the result of Ref. [5], I will give bounds for the overall rms noise level below which the tomographical reconstruction is still possible. I will show that the smearing effect of half photon of thermal noise in average is sufficient to completely spoil the measurement, making the experimental errors growing up unbounded.

2. Short Up-to-Date Review on Homodyne Tomography

The homodyne tomography method is designed to obtain a general matrix element $\langle \psi | \hat{\rho} | \phi \rangle$ in form of expectation of a function of the homodyne
outcomes at different phases with respect to the LO. In equations, one has
\[ \langle \psi | \hat{\varrho} | \varphi \rangle = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \, p(x; \phi) \, f_{\psi\varphi}(x; \phi), \] (1)
where \( p(x; \phi) \) denotes the probability distribution of the outcome \( x \) of the quadrature \( \hat{x}_\phi = \frac{1}{2} (a^\dagger e^{i\phi} + ae^{-i\phi}) \) of the field mode with particle operators \( a \) and \( a^\dagger \) at phase \( \phi \) with respect to the LO. Notice that it is sufficient to average only over \( \phi \in [0, \pi] \), due to the symmetry \( \hat{x}_{\phi+\pi} = -\hat{x}_\phi \). One wants the function \( f_{\psi\varphi}(x; \phi) \) bounded for all \( x \), whence every moment will be bounded for any possible (a priori unknown) probability distribution \( p(x; \phi) \). Then, according to the central-limit theorem, one is guaranteed that the integral in Eq. (1) can be sampled statistically over a sufficiently large set of data, and the average values for different experiments will be Gaussian distributed, allowing estimation of confidence intervals. If, on the other hand, the kernel \( f_{\psi\varphi}(x; \phi) \) turns out to be unbounded, then we will say that the matrix element cannot be measured by homodyne tomography.

The easiest way to obtain the integral kernel \( f_{\psi\varphi}(x; \phi) \) is starting from the operator identity
\[ \hat{\varrho} = \int \frac{d^2\alpha}{\pi} \Tr(\hat{\varrho} e^{-\alpha a + \alpha a^\dagger}) e^{-\alpha a^\dagger + \alpha a} \] (2)
which, by changing to polar variables \( \alpha = (i/2)ke^{i\phi} \), becomes
\[ \hat{\varrho} = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} \frac{dk}{4} |k| e^{ik\hat{x}_\phi} \] (3)
Equation (2) is nothing but the operator form of the Fourier-transform relation between Wigner function and characteristic function: it can also be considered as an operator form of the Moyal identity
\[ \int \frac{d^2z}{\pi} \langle k | \hat{D}^\dagger(z) | m \rangle \langle l | \hat{D}(z) | n \rangle = \langle k | n \rangle \langle l | m \rangle. \] (4)
The trace-average in Eq. (3) can be evaluated in terms of \( p(x, \phi) \), using the complete set \( \{|x\rangle_\phi \} \) of eigenvectors of \( \hat{x}_\phi \), and exchanging the integrals over \( x \) and \( k \). One obtains
\[ \hat{\varrho} = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \, p(x; \phi) K(x - \hat{x}_\phi), \] (5)
where the integral kernel \( K(x) \) is given by
\[ K(x) = -\frac{1}{2} P \frac{1}{x^2} \equiv -\lim_{\varepsilon \to 0^+} \frac{1}{2} \Re \frac{1}{(x + i\varepsilon)^2}, \] (6)
P denoting the Cauchy principal value. Taking matrix elements of both sides of Eq. (5) between vectors \( \psi \) and \( \varphi \), we obtain the sampling formula we were looking for, namely

\[
\langle \psi | \hat{\varrho} | \varphi \rangle = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \ p(x; \phi) \langle \psi | K(x - \hat{x}) | \varphi \rangle.
\]

(7)

Hence, the matrix element \( \langle \psi | \hat{\varrho} | \varphi \rangle \) is obtained by averaging the function \( f_{\psi,\varphi}(x; \phi) \equiv \langle \psi | K(x - \hat{x}) | \varphi \rangle \) over homodyne data at different phases \( \phi \). As we will see soon, despite \( K(x) \) is unbounded, for particular vectors \( \psi \) and \( \varphi \) in the Hilbert space the matrix element \( \langle \psi | K(x - \hat{x}) | \varphi \rangle \) is bounded, and thus the integral (7) can be sampled experimentally.

Before analyzing specific matrix representations, I recall how the sampling formula (7) can be generalized to the case of nonunit quantum efficiency. Low efficiency homodyne detection simply produces a probability \( p_\eta(x; \phi) \) that is a Gaussian convolution of the ideal probability \( p(x; \phi) \) for \( \eta = 1 \) (see, for example, Ref. [3]). In terms of the generating functions of the \( \hat{x} \)-moments one has

\[
\int_{-\infty}^{+\infty} \ dx \ p_\eta(x; \phi) e^{ikx} = \exp \left( -\frac{1 - \eta k^2}{8\eta} \right) \int_{-\infty}^{+\infty} \ dx \ p(x; \phi) e^{ikx}.
\]

(8)

Upon substituting Eq. (8) into Eq. (3), and by following the same lines that lead us to Eq. (5), one obtains the operator identity

\[
\hat{\varrho} = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \ p_\eta(x; \phi) K_\eta(x - \hat{x}),
\]

(9)

where now the kernel reads

\[
K_\eta(x) = \frac{1}{2} \text{Re} \int_{0}^{+\infty} \ dk \ k \ exp \left( \frac{1 - \eta k^2 + ikx}{8\eta} \right).
\]

(10)

The desired sampling formula for \( \langle \psi | \hat{\varrho} | \varphi \rangle \) is obtained again as in Eq. (7), by taking matrix elements of both sides of Eq. (10). Notice that now the kernel \( K_\eta(x) \) is not even a tempered distribution: however, as we will see immediately, the matrix elements of \( K_\eta(x - \hat{x}) \) are bounded for some representations, depending on the value of \( \eta \). The matrix elements \( \langle \psi | K_\eta(x - \hat{x}) | \varphi \rangle \) are bounded if the following inequality is satisfied for all phases \( \phi \in [0, \pi] \)

\[
\eta > \frac{1}{1 + 4\varepsilon^2(\phi)},
\]

(11)

where \( \varepsilon^2(\phi) \) is the harmonic mean

\[
\frac{2}{\varepsilon^2(\phi)} = \frac{1}{\varepsilon^2_\psi(\phi)} + \frac{1}{\varepsilon^2_\varphi(\phi)},
\]

(12)
and $\varepsilon^2_v(\phi)$ is the “resolution” of the vector $|v\rangle$ in the $\hat{x}_\phi$-representation, namely:

$$|\phi\langle x|v\rangle|^2 \simeq \exp\left[-\frac{x^2}{2\varepsilon^2_v(\phi)}\right]. \quad (13)$$

In Eq. (13) the symbol $\simeq$ stands for the leading term as a function of $x$, and $|x\rangle_\phi \equiv e^{i a^\dagger a^\phi}|x\rangle$ denote eigen-ket of the quadrature $\hat{x}_\phi$ for eigenvalue $x$. Upon maximizing Eq. (11) with respect to $\phi$ one obtains the bound

$$\eta > \frac{1}{1 + 4\varepsilon^2}, \quad \varepsilon^2 = \min_{\phi\in[0,\pi]} \{\varepsilon^2(\phi)\}. \quad (14)$$

One can easily see that the bound is $\eta > 1/2$ for both number-state and coherent-state representations, whereas it is $\eta > (1 + s^2)^{-1} \geq 1/2$ for squeezed-state representations with minimum squeezing factor $s < 1$. On the other hand, for the quadrature representation one has $\eta > 1$, which means that this matrix representation cannot be measured. The value $\eta = 1/2$ is actually an absolute bound for all representations satisfying the “Heisenberg relation” $\varepsilon(\phi)\varepsilon(\phi + \frac{\pi}{2}) \geq \frac{1}{4}$ with the equal sign, which include all known representations (for a discussion on the existence of exotic representations see Ref. [4]). Here, I want to emphasize that the existence of such a lower bound for quantum efficiency is actually of fundamental relevance, as it prevents measuring the wave function of a single system using schemes of weak repeated indirect measurements on the same system [10].

At the end of this section, from Ref. [5] I report for completeness the kernel $\langle n|K(x - \hat{x}_\phi)|m\rangle$ for matrix elements between number eigenstates. One has

$$\langle n|K_\eta(x - \hat{x}_\phi)|n + d\rangle = e^{-id\phi}2\kappa^{d+2}\sqrt{n!}\frac{n!}{(n+d)!}e^{-\kappa^2x^2}$$

$$\times \sum_{\nu=0}^{n} \frac{(-)^\nu}{\nu!} \binom{n+d}{n-\nu} (2\nu + d + 1)!\kappa^{2\nu}\text{Re} \left\{(-i)^d D_{-(2\nu+d+2)}(-2i\kappa x)\right\}, \quad (15)$$

where $\kappa = \sqrt{\eta/(2\eta - 1)}$, and $D_{\sigma}(z)$ denotes the parabolic cylinder function. For $\eta = 1$ the kernel factorizes as follows [19, 14]

$$\langle n|K(x - \hat{x}_\phi)|n + d\rangle$$

$$= e^{-id\phi}\left[2xu_n(x)v_{n+d}(x) - \sqrt{n+1}u_{n+1}(x)v_{n+d}(x) - \sqrt{m+1}u_n(x)v_{n+d+1}(x)\right], \quad (16)$$
where $u_n(x)$ and $v_n(x)$ are the regular and irregular energy eigen-functions of the harmonic oscillator

$$u_j(x) = \frac{1}{\sqrt{j!}} \left( x - \frac{\partial_x}{2} \right)^j \left( \frac{2}{\pi} \right)^{1/4} e^{-x^2},$$

$$v_j(x) = \frac{1}{\sqrt{j!}} \left( x - \frac{\partial_x}{2} \right)^j (2\pi)^{1/4} e^{-x^2} \int_0^{\sqrt{2}x} dt \, e^{t^2}. \quad (17)$$

3. Measuring Generic Field Operators

Homodyne tomography provides the maximum achievable information on the quantum state, and, in principle, the knowledge of the density matrix should allow one to calculate the expectation value $\langle \hat{O} \rangle = \text{Tr}[\hat{O} \hat{\varrho}]$ of any observable $\hat{O}$. However, this is generally true only when one has an analytic knowledge of the density matrix, but it is not true when the matrix has been obtained experimentally. In fact, the Hilbert space is actually infinite dimensional, whereas experimentally one can achieve only a finite matrix, each element being affected by an experimental error. Notice that, even though the method allows one to extract any matrix element in the Hilbert space from the same bunch of experimental data, however, it is the way in which errors converge in the Hilbert space that determines the actual possibility of estimating the trace $\text{Tr}[\hat{O} \hat{\varrho}]$. To make things more concrete, let us fix the case of the number representation, and suppose we want to estimate the average photon number $\langle a^\dagger a \rangle$. In Ref. [9] it has been shown that for nonunit quantum efficiency the statistical error for the diagonal matrix element $\langle n | \hat{\varrho} | n \rangle$ diverges faster than exponentially versus $n$, whereas for $\eta = 1$ the error saturates for large $n$ to the universal value $\varepsilon_n = \sqrt{2/N}$ that depends only on the number $N$ of experimental data, but is independent on both $n$ and on the quantum state. Even for the unrealistic case $\eta = 1$, one can see immediately that the estimated expectation value $\langle a^\dagger a \rangle = \sum_{n=0}^{H-1} n \varrho_{nn}$ based on the measured matrix elements $\varrho_{nn}$, is not guaranteed to converge versus the truncated-space dimension $H$, because the error on $\varrho_{nn}$ is nonvanishing versus $n$. Clearly in this way I am not proving that the expectation $\langle a^\dagger a \rangle$ is unobtainable from homodyne data, because matrix errors convergence depends on the chosen representation basis, whence the ineffectiveness of the method may rely in the data processing, more than in the actual information contained in the bunch of experimental data. Therefore, the question is: is it possible to estimate a generic expectation value $\langle \hat{O} \rangle$ directly from homodyne data, without using the measured density matrix? As we will see soon, the answer is positive in most cases of interest,
and the procedure for estimating the expectation $\langle \hat{O} \rangle$ will be referred to as "homodyning the observable $\hat{O}$.

By "homodyning the observable $\hat{O}$" I mean averaging an appropriate kernel function $\mathcal{R}[\hat{O}](x; \phi)$ (independent on the state $\hat{\varrho}$) over the experimental homodyne data, achieving in this way the expectation value of the observable $\langle \hat{O} \rangle$ for every state $\hat{\varrho}$. Hence, the kernel function $\mathcal{R}[\hat{O}](x; \phi)$ is defined through the identity

$$\langle \hat{O} \rangle = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \ p(x; \phi) \mathcal{R}[\hat{O}](x; \phi).$$

(18)

From the definition of $\mathcal{R}[\hat{O}](x; \phi)$ in Eq. (18), and from Eqs. (2) and (3) — which generally hold true for any Hilbert-Schmidt operator in place of $\hat{\varrho}$ — one obtains

$$\hat{O} = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \mathcal{R}[\hat{O}](x; \phi) |x\rangle \langle x|,$$

(19)

with the kernel $\mathcal{R}[\hat{O}](x; \phi)$ given by

$$\mathcal{R}[\hat{O}](x; \phi) = \text{Tr}[\hat{O} \mathcal{K}(x - \hat{x})],$$

(20)

and $\mathcal{K}(x)$ given in Eq. (6). The validity of Eq. (20), however, is limited only to the case of a Hilbert-Schmidt operator $\hat{O}$, otherwise it is ill defined. Nevertheless, one can obtain the explicit form of the kernel $\mathcal{R}[\hat{O}](x; \phi)$ in a different way. Starting from the identity involving trilinear products of Hermite polynomials [11]

$$\int_{-\infty}^{+\infty} dx \ e^{-x^2} H_k(x) H_m(x) H_n(x) = \frac{2^{\frac{m+n+k}{2}} \pi^{\frac{1}{2}} k! m! n!}{(s-k)!(s-m)!(s-n)!},$$

(21)

for $k + m + n = 2s$ even.

Richter proved the following nontrivial formula for the expectation value of the normally ordered field operators [20]

$$\langle a^\dagger^n a^m \rangle = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \ p(x; \phi) e^{i(m-n)\phi} \frac{H_{n+m}(\sqrt{2}x)}{\sqrt{2^{n+m}n^m}},$$

(22)

which corresponds to the kernel

$$\mathcal{R}[a^\dagger^n a^m](x; \phi) = e^{i(m-n)\phi} \frac{H_{n+m}(\sqrt{2}x)}{\sqrt{2^{n+m}n^m}}.$$

(23)

This result can be easily extended to the case of nonunit quantum efficiency $\eta < 1$, as the normally ordered expectation $\langle a^\dagger^n a^m \rangle$ just gets an extra factor
\( \eta^{2(n+m)} \). Therefore, one has

\[
\mathcal{R}_\eta[a^\dagger a^m](x; \phi) = e^{i(m-n)\phi} \frac{H_{a+m}(\sqrt{x})}{\sqrt{(2\eta)^{n+m}(n+m)}} ,
\]

(24)

where the kernel \( \mathcal{R}_\eta[\hat{O}](x; \phi) \) is defined as in Eq. (18), but with the experimental probability distribution \( p_\eta(x; \phi) \). From Eq. (24) by linearity on can obtain the kernel \( \mathcal{R}_\eta[\hat{f}](x; \phi) \) for any operator function \( \hat{f} \) that has normal ordered expansion

\[
\hat{f} \equiv f(a, a^\dagger) = \sum_{nm=0}^{\infty} f_{nm} a^\dagger a^m .
\]

(25)

From Eq. (24) one obtains

\[
\mathcal{R}_\eta[\hat{f}](x; \phi) = \sum_{s=0}^{\infty} \frac{H_s(\sqrt{x})}{s!(2\eta)^{s/2}} \sum_{nm=0}^{\infty} f_{nm} e^{i(m-n)\phi} n!m! \delta_{n+m,s}
\]

\[
= \sum_{s=0}^{\infty} \frac{H_s(\sqrt{x})^s}{s!(2\eta)^{s/2}} \int \frac{d^s}{dv^s} F[\hat{f}](v; \phi),
\]

(26)

where

\[
F[\hat{f}](v; \phi) = \sum_{nm=0}^{\infty} f_{nm} \left( \frac{n+m}{m} \right)^{-1} (-iv)^{n+m} e^{i(m-n)\phi} .
\]

(27)

Continuing from Eq. (26) one obtains

\[
\mathcal{R}_\eta[\hat{f}](x; \phi) = \exp \left( \frac{1}{2\eta} \frac{d^2}{dv^2} + \frac{2ix}{\sqrt{\eta}} d \right) F[\hat{f}](v; \phi) ,
\]

(28)

and finally

\[
\mathcal{R}_\eta[\hat{f}](x; \phi) = \int_{-\infty}^{+\infty} \frac{dw}{\sqrt{2\pi\eta^{-1}}} e^{-\frac{4}{\eta}w^2} F[\hat{f}](w + 2ix/\sqrt{\eta}; \phi) .
\]

(29)

Hence one concludes that the operator \( \hat{f} \) can be measured by homodyne tomography if the function \( F[\hat{f}](v; \phi) \) in Eq. (27) grows slower than \( \exp(-\eta v^2/2) \) for \( v \to \infty \), and the integral in Eq. (29) grows at most exponentially for \( x \to \infty \) (assuming \( p(x; \phi) \) goes to zero faster than exponentially at \( x \to \infty \)).

In Table 3 I report the kernel \( \mathcal{R}_\eta[\hat{O}](x; \phi) \) for some operators \( \hat{O} \). One can see that for the raising operator \( \hat{c}_+ \), the kernel diverges at \( \eta = 1/2^+ \), namely it can be measured only for \( \eta > 1/2 \). The operator \( \hat{W}_s \) in the same table gives the generalized Wigner function \( W_s(\alpha, \bar{\alpha}) \) for ordering parameter \( s \).
through the identity $W_s(\alpha, \tilde{\alpha}) = \text{Tr}[\hat{D}(\alpha) \hat{D}^\dagger(\tilde{\alpha}) \hat{W}_s]$. From the expression of $\mathcal{R}_\eta[\hat{W}_s](x; \phi)$ it follows that by homodyning with quantum efficiency $\eta$ one can measure the generalized Wigner function only for $s < 1 - \eta^{-1}$; in particular, as already noticed in Refs. [5], the usual Wigner function for $s = 0$ cannot be measured for any quantum efficiency [in fact one would have $\mathcal{R}_1[\hat{D}^\dagger(\alpha) \hat{W}_0 \hat{D}(\alpha)](x; \phi) = K[x - \text{Re}(\alpha e^{-i\phi})]$, with $K(x)$ unbounded as given in Eq. (6)].

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
& $\hat{O}$ & $\mathcal{R}_\eta[\hat{O}](x; \phi)$ \\
\hline
(1) & $a^{\dagger n} a^m$ & $e^{i(m-n)\phi} \frac{H_{n+m}(\sqrt{2}x)}{\sqrt{2^{n+m} n! m!}}$ \\
\hline
(2) & $a$ & $2e^{i\phi} x$ \\
\hline
(3) & $a^2$ & $e^{2i\phi}(4x^2 - 1)$ \\
\hline
(4) & $a^{\dagger} a$ & $2x^2 - \frac{1}{2}$ \\
\hline
(5) & $(a^{\dagger} a)^2$ & $\frac{8}{3} x^4 - 2x^2$ \\
\hline
(6) & $\hat{D}^\dagger(\alpha) := e^{-\alpha a^\dagger} e^{\alpha a}$ & $\frac{\exp[-\frac{1}{4\eta}(\alpha e^{i\phi})^2 + \frac{\pi}{2\eta} i\alpha e^{i\phi}]}{1 + \frac{\pi}{4\eta} e^{-2i\phi}} + \frac{\exp[-\frac{1}{4\eta}(\alpha e^{-i\phi})^2 - \frac{\pi}{2\eta} i\alpha e^{-i\phi}]}{1 + \frac{\pi}{4\eta} e^{2i\phi}}$ \\
\hline
(7) & $\hat{e}_+ := a^{\dagger} \frac{1}{\sqrt{1 + a^{\dagger} a}}$ & $2xe^{-i\phi} \frac{1}{\sqrt{2\pi\eta}} \int_{-\infty}^{+\infty} dv \Phi \left(2, 3; \frac{1 + z^2}{1 + z - 1}; \frac{2}{2\eta^2} \right)$, \\
& & $z = \frac{e^{-\frac{1}{2\eta}}}{e^{-\frac{1}{2\eta}}}$ \\
\hline
(8) & $\hat{W}_s \triangleq \frac{2}{\pi(1-s)} \left( \frac{s+1}{s-1} \right)^{a^{\dagger} a}$ & $\int_0^\infty dt \frac{2e^{-t}}{\pi \left(1 - s - \frac{1}{\eta}\right)} \cos \left(2 \sqrt{\frac{2t}{(1-s) - \frac{1}{\eta}}} \right)$ \\
\hline
(9) & $|n + d\rangle \langle n|$ & $\langle n | K(x - \hat{x}_\phi) | n + d\rangle$ in Eqs. (15) and (16) \\
\hline
\end{tabular}
\caption{Kernel $\mathcal{R}_\eta[\hat{O}](x; \phi)$, as defined in Eq. (18), for some operators $\hat{O}$. [The symbol $\Phi(a, b; x)$ denotes the customary confluent hypergeometric function.]
}
\end{table}

### 3.1. Comparison between homodyne tomography and heterodyning

We have seen that from the same bunch of homodyne tomography data, not only one can recover the density matrix of the field, but also one can measure any field observable $\hat{f} \equiv f(a, a^\dagger)$ having normal ordered expansion $\hat{f} \equiv f^{(n)}(a, a^\dagger) = \sum_{nm=0}^{\infty} f^{(n)}_{nm} a^{\dagger n} a^m$ and bounded integral in Eq. (29) — this holds true in particular for any polynomial function
of the annihilation and creation operators. This situation can be compared with the case of heterodyne detection, where again one measures general field observables, but admitting anti-normal ordered expansion

$$\hat{f} \equiv f^{(a)}(a, a^\dagger) = \sum_{nm=0}^{\infty} f^{(a)}_{nm} a^m a^{\dagger n},$$

in which case the expectation value is obtained through the heterodyne average

$$\langle \hat{f} \rangle = \int \frac{d^2 \alpha}{\pi} f^{(a)}(\alpha, \overline{\alpha}) \langle \alpha | \hat{\rho} | \alpha \rangle.$$  \hspace{1cm} (30)

For $\eta = 1$ the heterodyne probability is just the $Q$-function $Q(\alpha, \overline{\alpha}) = \frac{1}{\pi} \langle \alpha | \hat{\rho} | \alpha \rangle$, whereas for $\eta = 1$ it will be Gaussian convoluted. As shown by Baltin [1], generally the anti-normal expansion either is not defined, or is not consistent on the Fock basis, namely $f^{(a)}(a, a^\dagger)|n\rangle$ has infinite norm or is different from $\hat{f}(a, a^\dagger)|n\rangle$ for some $n \geq 0$. In particular, let us focus attention on functions of the number operator $f(a^\dagger a) = \sum_{l=0}^{\infty} c_l (a^\dagger a)^l$, $f^{(n)}(a^\dagger a) = \sum_{l=0}^{\infty} c^{(n)}_l a^l a^\dagger$, $f^{(a)}(a^\dagger a) = \sum_{l=0}^{\infty} c^{(a)}_l a^l a^\dagger$. Baltin has shown that [1]

$$c^{(n)}_l = \frac{1}{l!} \int_{-\infty}^{+\infty} d\lambda g(\lambda)(e^{-i\lambda} - 1)^l = \sum_{k=0}^{l} \frac{(-)^{l-k} f(k)}{k!(l-k)!},$$

$$c^{(a)}_l = \frac{1}{l!} \int_{-\infty}^{+\infty} d\lambda e^{i\lambda} g(\lambda)(1 - e^{i\lambda})^l = \sum_{k=0}^{l} \frac{(-)^k f(-k - 1)}{k!(l-k)!},$$ \hspace{1cm} (31)

$$g(\lambda) = \int_{-\infty}^{+\infty} \frac{dx}{2\pi} f(x)e^{i\lambda x}.$$  

From Eqs. (31) one can see that the normal ordered expansion is always well defined, whereas the anti-normal ordering needs extending the domain of $f$ to negative integers. However, even though the anti-normal expansion is defined, this does not mean that the expectation of $f(a^\dagger a)$ can be obtained through heterodyning, because the integral in Eq. (30) may not exist. Actually, this is the case when the anti-normal expansion is not consistent on the Fock basis. In fact, for the exponential function $f(a^\dagger a) = \exp(-\mu a^\dagger a)$ one has $f^{(a)}(|\alpha|^2) = e^{\mu} \exp[(1 - e^{\mu})|\alpha|^2]$; on the Fock basis $f^{(a)}(a^\dagger a)|n\rangle$ is a binomial expansion with finite convergence radius, and this gives the consistency condition $|1 - e^{\mu}| < 1$. However, one can take the analytic continuation corresponding for $1 - e^{\mu} < 1$, which coincides with the condition that the integral in Eq. (30) exists for any state $\hat{\rho}$ (the $Q$-function vanishes as $\exp(-|\alpha|^2)$ for $\alpha \rightarrow \infty$, at least for states with limited photon number). This argument can be extended by Fourier transform to more general functions $f(a^\dagger a)$, leading to the conclusion that there are field op-
dom variable erators, and compare them with the heterodyne noise. For a complex ran-
the random variable is $z \equiv 2e^{i\phi}x$ [22] and the average over-line denotes the
double integral over $x$ and $\phi$ in Eq. (18). From Table (3) one has $\overline{\varphi} = \langle a \rangle$, $\overline{z^2} = \langle a^2 \rangle$, $\langle z \rangle^2 = 2\langle a \rangle + 1$, $e^{2i\varphi} = 0$ [23]. In this way one finds that the noise from homodyning the field is $N_{\text{hom}}^{(\pm)}[a] = 1 + 2\langle a \rangle - \langle a \rangle^2 \pm \langle a^2 \rangle - \langle a \rangle^2$.
On the other hand, when heterodyning, $z$ becomes the heterodyne output photocurrent, whence $\overline{\varphi} = \langle a \rangle$, $\overline{z^2} = \langle a^2 \rangle$, $\langle z \rangle^2 = \langle a \rangle + 1$, and one has $N_{\text{het}}^{(\pm)}[a] = 1 + \langle a \rangle - \langle a \rangle^2 \pm \langle a^2 \rangle - \langle a \rangle^2$, so that the tomographical noise is larger than the heterodyne noise by a term equal to the average photon number, i. e.

$$N_{\text{hom}}^{(\pm)}[a] = N_{\text{het}}^{(\pm)}[a] + \langle a \rangle$$  \hspace{1cm} (32)

Therefore, homodyning the field is always more noisy than heterodyning it. On the other hand, for other field observables it may happen that homodyne tomography is less noisy than heterodyne detection. For example, one can easily evaluate the noise $N_{\text{hom}}[\hat{n}]$ when homodyning the photon number $\hat{n} = a^\dagger a$. The random variable corresponding to the photon number is $\nu(z) = \frac{1}{2}(\langle z \rangle^2 - 1) \equiv 2x^2 - \frac{1}{2}$, and from Table 3 we see that the noise $N_{\text{hom}}[\hat{n}] = \Delta\nu^2(z)$ can be written as $N_{\text{hom}}[\hat{n}] = \langle \Delta\hat{n}^2 \rangle + \frac{1}{2}\langle \hat{n}^2 + \hat{n} + 1 \rangle$ [9]. When heterodyning the field, the random variable corresponding to the photon number is $\nu(z) = |z|^2 - 1$, and from the relation $|z|^4 = \langle a^\dagger a \rangle$ one
obtains $N_{\text{het}}[\hat{n}] \doteq \overline{\Delta \nu^2(z)} = \langle \Delta \hat{n}^2 \rangle + \langle \hat{n} + 1 \rangle$, namely

$$N_{\text{hom}}[\hat{n}] = N_{\text{het}}[\hat{n}] + \frac{1}{2} \langle \hat{n}^2 - \hat{n} - 1 \rangle.$$  \hspace{1cm} (33)

We thus conclude that homodyning the photon number is less noisy than heterodyning it for sufficiently low mean photon number $\langle \hat{n} \rangle < \frac{1}{2}(1 + \sqrt{5})$.

4. Homodyne Tomography in Presence of Additive Phase-Insensitive Noise

In this section I consider the case of additive Gaussian noise, in the typical situation in which the noise is phase-insensitive. This kind of noise is described by a density matrix evolved by the master equation

$$\partial_t \hat{\rho}(t) = 2 \left[ AL[\hat{a}^\dagger] + BL[\hat{a}] \right] \hat{\rho}(t),$$  \hspace{1cm} (34)

where $L[\hat{c}]$ denotes the Lindblad super-operator $L[\hat{c}] \hat{\rho} = \hat{c} \hat{\rho} \hat{c}^\dagger - \frac{1}{2} [\hat{c}^\dagger \hat{c}, \hat{\rho}]_+$. Due to the phase invariance $L[ae^{-i\phi}] = L[a]$ the dynamical evolution does not depend on the phase, and the noise is phase insensitive. From the evolution of the averaged field $\langle a \rangle_{\text{out}} \equiv \text{Tr}[\hat{a} \hat{\rho}(t)] = g \langle a \rangle_{\text{in}} \equiv \text{Tr}[\hat{a} \hat{\rho}(0)]$ with $g = \exp[(A - B)t]$, we can see that for $A > B$ Eq. (34) describes phase-insensitive amplification with field-gain $g$, whereas for $B > A$ it describes phase-insensitive attenuation, with $g < 1$. Concretely, for $A > B$ Eq. (34) models unsaturated parametric amplification with thermal idler [average photon number $\bar{m} = B/(A - B)$], or unsaturated laser action [$A$ and $B$ proportional to atomic populations on the upper and lower lasing levels respectively]. For $B > A$, on the other hand, the same equation describes a field mode damped toward the thermal distribution [inverse photon lifetime $\Gamma = 2(B - A)$], equilibrium photon number $\bar{m} = A/(B - A)$], or a loss $g < 1$ along an optical fiber or at a beam-splitter, or even due to frequency conversion [6]. The borderline case $A = B$ leaves the average field invariant, but introduces noise that changes the average photon number as $\langle a^\dagger a \rangle_{\text{out}} = \langle a^\dagger a \rangle_{\text{in}} + \bar{n}$, where $\bar{n} = 2At$. In this case the solution of Eq. (34) can be cast into the simple form

$$\hat{\rho}(t) = \int \frac{d^2\beta}{\pi \bar{n}} \exp(-|\beta|^2/\bar{n}) \hat{D}(\beta) \hat{\rho}(0) \hat{D}^\dagger(\beta).$$  \hspace{1cm} (35)

This is the Gaussian displacement noise studied in Refs. [13, 12] and commonly referred to as “thermal noise” [regarding the misuse of this terminology, see Ref. [12]], which can be used to model many kinds of undesired environmental effects, typically due to linear interactions with random classical fluctuating fields.
Eq. (34) has the following simple Fokker-Planck differential representation [8] in terms of the generalized Wigner function $W_s(\alpha, \tilde{\alpha})$ for ordering parameter $s$

$$\partial_t W_s(\alpha, \tilde{\alpha}; t) = [Q(\partial_\alpha \alpha + \partial_{\tilde{\alpha}} \tilde{\alpha}) + 2 D_s \partial^2_{\alpha, \tilde{\alpha}}] W_s(\alpha, \tilde{\alpha}; t) ,$$  \hspace{1cm} (36)

where $Q = B - A$ and $2 D_s = A + B + s(A - B)$. For nonunit quantum efficiency $\eta$ and after a noise-diffusion time $t$ the homodyne probability distribution $p_\eta(x; \phi; t)$ can be evaluated as the marginal distribution of the Wigner function for ordering parameter $s = 1 - \eta^{-1}$, namely

$$p_\eta(x; \phi; t) = \int_{-\infty}^{+\infty} dy W_{1-\eta^{-1}} ((x + iy)e^{i\phi}, (x - iy)e^{-i\phi}; t) .$$  \hspace{1cm} (37)

The solution of Eq. (36) is the Gaussian convolution [8]

$$W_s(\alpha, \tilde{\alpha}; t) = \int \frac{d^2 \beta}{\pi \delta^2_s} \exp \left[ -\frac{|\alpha - g\beta|^2}{\delta^2_s} \right] W_s(\beta, \bar{\beta}; 0) ,$$

$$\delta^2_s = \frac{D_s}{Q} (1 - e^{-2Qt}) ,$$  \hspace{1cm} (38)

and using Eq. (37) one obtains the homodyne probability distribution

$$p_\eta(x; \phi; t) = e^{Qt} \int_{-\infty}^{\infty} dx' \frac{dx'}{\sqrt{2\pi \Delta^2_1 - \eta^{-1}}} \exp \left[ -\frac{(x' - g^{-1} x)^2}{2\Delta^2_1 - \eta^{-1}} \right] p_\eta(x'; \phi) ,$$  \hspace{1cm} (39)

where $\Delta^2_\eta = \frac{1}{2} g^2 \delta^2_{1-\eta^{-1}}$. It is easy to see that the generating function of the $\hat{x}_\phi$-moments with the experimental probability $p_\eta(x; \phi; t)$ can be written in terms of the probability distribution $p(x; \phi)$ for perfect homodyning as follows

$$\int_{-\infty}^{+\infty} dx \, p_\eta(x; \phi; t)e^{ikx} =$$

$$\exp \left( -\frac{1}{2} g^2 \Delta^2_\eta k^2 - \frac{1}{8\eta} g^2 k^2 \right) \int_{-\infty}^{+\infty} dx \, p(x; \phi)e^{ikx} .$$  \hspace{1cm} (40)

Eq. (40) has the same form of Eq. (8), but with the Fourier variable $k$ multiplied by $g$ and with an overall effective quantum efficiency $\eta_*$ given by

$$\eta_*^{-1} = \eta^{-1} + 4 \Delta^2_\eta = g^{-2}\eta^{-1} + \frac{2A}{B - A}(g^{-2} - 1) .$$  \hspace{1cm} (41)

On the other hand, following the same lines that lead us to Eq. (9), we obtain the operator identity

$$\hat{\phi} \equiv \hat{\phi}(0) = \int_0^\pi \frac{d\phi}{\pi} \int_{-\infty}^{+\infty} dx \, p_{\eta_*}(x; \phi; t) K_{\eta_*}(g^{-1} x - \hat{x}_\phi) ,$$  \hspace{1cm} (42)
which also means that when homodyning the operator $\hat{O}$ one should use $\mathcal{R}_{\eta}(g^{-1}x; \phi)$ in place of $\mathcal{R}_{\eta}(x; \phi)$, namely, more generally, one needs to re-scale the homodyne outcomes by the gain and use the effective quantum efficiency $\eta_*$ in Eq. (41). In terms of the gain $g$ and of the input-output photon numbers, the effective quantum efficiency reads

$$\eta_*^{-1} = \eta^{-1} + g^{-2}(2 \langle a^\dagger a \rangle_{\text{out}} + \eta^{-1}) - (2 \langle a^\dagger a \rangle_{\text{in}} + \eta^{-1}).$$

(43)

In the case of pure displacement Gaussian noise ($A = B$), Eq. (43) becomes

$$\eta_*^{-1} = \eta^{-1} + 2\bar{n},$$

(44)

which means that the bound $\eta_* > 1/2$ is surpassed already for $\bar{n} \geq 1$: in other worlds, it is just sufficient to have half photon of thermal noise to completely spoil the tomographic reconstruction.

References


[22] Notice that for the complex random variable \( z = 2e^{i\phi}x \) the phase \( \phi \) is a scanning parameter imposed by the detector. (Actually, the best way to experimentally scan the integral in Eq. (18) is just to pick up the phase \( \phi \) at random.) Nevertheless, the argument of the complex number \( z \) is still a genuine random variable, because the sign of \( x \) is random, and depends on the value of \( \phi \). One has \( \text{arg}(z) = \phi + \pi(1 - \text{sgn}(x)) \). For example, for any highly excited coherent state \( |\alpha\rangle \) the probability distribution of \( \text{arg}(z) \) will approach a uniform distribution on \( [\text{arg}(\alpha) - \pi/2, \text{arg}(\alpha) + \pi/2] \).

[23] One should remember that, the phase \( \phi \) is imposed by the detector, and is uniformly scanned (randomly or not) in the interval \([0, \pi]\). This leads to \( e^{2i\phi} = 0 \), independently on the state \( \hat{\phi} \).
CHAPTER 34

On the Measurement of Qubits

Daniel F. V. James, Paul G. Kwiat, William J. Munro, and Andrew G. White

Abstract. We describe in detail the theory underpinning the measurement of density matrices of a pair of quantum two-level systems ("qubits"). Our particular emphasis is on qubits realized by the two polarization degrees of freedom of a pair of entangled photons generated in a down-conversion experiment; however, the discussion applies in general, regardless of the actual physical realization. Two techniques are discussed, namely, a tomographic reconstruction (in which the density matrix is linearly related to a set of measured quantities) and a maximum likelihood technique which requires numerical optimization (but has the advantage of producing density matrices which are always non-negative definite). In addition, a detailed error analysis is presented, allowing errors in quantities derived from the density matrix, such as the entropy or entanglement of formation, to be estimated. Examples based on down-conversion experiments are used to illustrate our results.

1. Introduction

The ability to create, manipulate, and characterize quantum states is becoming an increasingly important area of physical research, with implications for areas of technology such as quantum computing, quantum cryptography, and communications. With a series of measurements on a large enough number of identically prepared copies of a quantum system, one can infer, to a reasonable approximation, the quantum state of the system. Arguably, the first such experimental technique for determining the state of a quantum system was devised by George Stokes in 1852 [23]. His famous four parameters allow an experimenter to determine uniquely the polarization state of a light beam. With the insight provided by nearly 150 years of progress in optical physics, we can consider coherent light beams to be an
ensemble of two-level quantum mechanical systems, the two levels being the
two polarization degrees of freedom of the photons; the Stokes parameters
allow one to determine the density matrix describing this ensemble. More
recently, experimental techniques for the measurement of the more subtle
quantum properties of light have been the subject of intensive investiga-
tion (see Ref. [15] for a comprehensive and erudite exposition of this sub-
ject). In various experimental circumstances it has been found reasonably
straightforward to devise a simple linear tomographic technique in which
the density matrix (or Wigner function) of a quantum state is found from a
linear transformation of experimental data. However, there is one important
drawback to this method, in that the recovered state might not correspond
to a physical state because of experimental noise. For example, density
matrices for any quantum state must be Hermitian, positive semi-definite
matrices with unit trace. The tomographically measured matrices often fail
to be positive semi-definite, especially when measuring low-entropy states.
To avoid this problem the “maximum likelihood” tomographic approach
to the estimation of quantum states has been developed [8, 24, 2, 9, 20]. In
this approach the density matrix that is “mostly likely” to have produced
a measured data set is determined by numerical optimization.

In the past decade several groups have successfully employed tomo-
graphic techniques for the measurement of quantum mechanical systems.
In 1990 Risley et al. reported the measurement of the density matrix for the
nine sublevels of the $n = 3$ level of hydrogen atoms formed following colli-
sion between $\text{H}^+$ ions and $\text{He}$ atoms, in conditions of high symmetry which
simplified the tomographic problem [1]. Since then, in 1993 Smithey et al.
made a homodyne measurement of the Wigner function of a single mode
of light [19]. Other explorations of the quantum states of single mode light
fields have been made by Mlynek et al. [11] and Bachor et al.[30]. Other
quantum systems whose density matrices have been investigated experi-
mentally include the vibrations of molecules [6], the motion ions and atoms
[14, 10], and the internal angular momentum quantum state of the $F = 4$
ground state of a cesium atom [12]. The quantum states of multiple spin-
$\frac{1}{2}$ nuclei have been measured in the high-temperature regime using NMR
techniques [5], albeit in systems of such high entropy that the creation of
entangled states is necessarily precluded [4]. The measurement of the quan-
tum state of entangled qubit pairs, realized using the polarization degrees
of freedom of a pair of photons created in a parametric down-conversion
experiment, was reported by us recently [26].

In this paper we will examine techniques in detail for quantum state
measurement as it applies to multiple correlated two-level quantum mechanical systems (or “qubits” in the terminology of quantum information). Our particular emphasis is qubits realized via the two polarization degrees of freedom of photons, data from which we use to illustrate our results. However, these techniques are readily applicable to other technologies proposed for creating entangled states of pairs of two-level systems. Because of the central importance of qubit systems to the emergent discipline of quantum computation, a thorough explanation of the techniques needed to characterize the qubit states will be of relevance to workers in the various diverse experimental fields currently under consideration for quantum computation technology [31]. This paper is organized as follows: In Section 2 we explore the analogy with the Stokes parameters, and how they lead naturally to a scheme for measurement of an arbitrary number of two-level systems. In Section 3, we discuss the measurement of a pair of qubits in more detail, presenting the validity condition for an arbitrary measurement scheme and introducing the set of 16 measurements employed in our experiments. Section 4 deals with our method for maximum likelihood reconstruction and in Section 5 we demonstrate how to calculate the errors in such measurements, and how these errors propagate to quantities calculated from the density matrix.

2. The Stokes Parameters and Quantum State Tomography

As mentioned above, there is a direct analogy between the measurement of the polarization state of a light beam and the measurement of the density matrix of an ensemble of two-level quantum mechanical systems. Here we explore this analogy in more detail.

2.1. Single qubit tomography

The Stokes parameters are defined from a set of four intensity measurements [7]: (i) with a filter that transmits 50% of the incident radiation, regardless of its polarization; (ii) with a polarizer that transmits only horizontally polarized light; (iii) with a polarizer that transmits only light polarized at 45° to the horizontal; and (iv) with a polarizer that transmits only right circularly polarized light. The number of photons counted by a detector, which is proportional to the classical intensity, in these four experiments
are as follows:
\[ n_0 = \frac{N}{2} \left( \langle H | \hat{\rho} | H \rangle + \langle V | \hat{\rho} | V \rangle \right) = \frac{N}{2} \left( \langle R | \hat{\rho} | R \rangle + \langle L | \hat{\rho} | L \rangle \right), \]
\[ n_1 = N \left( \langle H | \hat{\rho} | H \rangle \right) = \frac{N}{2} \left( \langle R | \hat{\rho} | R \rangle + \langle R | \hat{\rho} | L \rangle + \langle L | \hat{\rho} | R \rangle + \langle L | \hat{\rho} | L \rangle \right), \]
\[ n_2 = N \left( \langle \bar{D} | \hat{\rho} | \bar{D} \rangle \right) = \frac{N}{2} \left( \langle R | \hat{\rho} | R \rangle + \langle L | \hat{\rho} | L \rangle - i \langle L | \hat{\rho} | R \rangle + i \langle R | \hat{\rho} | L \rangle \right), \]
\[ n_3 = N \left( \langle R | \hat{\rho} | R \rangle \right). \]  

Here \( |H\rangle, |V\rangle, |\bar{D}\rangle = (|H\rangle - |V\rangle) / \sqrt{2} = \exp(i\pi/4) \left(|R\rangle + i|L\rangle\right) / \sqrt{2} \) and \( |R\rangle = (|H\rangle - i|V\rangle) / \sqrt{2} \) are the kets representing qubits polarized in the linear horizontal, linear vertical, linear diagonal (45°), and right-circular senses respectively, \( \hat{\rho} \) is the (2 × 2) density matrix for the polarization degrees of the light (or for a two-level quantum system), and \( N \) is a constant dependent on the detector efficiency and light intensity. The Stokes parameters, which fully characterize the polarization state of the light, are then defined by
\[ S_0 \equiv 2n_0 = N \left( \langle R | \hat{\rho} | R \rangle + \langle L | \hat{\rho} | L \rangle \right), \]
\[ S_1 \equiv 2 \left( n_1 - n_0 \right) = N \left( \langle R | \hat{\rho} | L \rangle + \langle L | \hat{\rho} | R \rangle \right), \]
\[ S_2 \equiv 2 \left( n_2 - n_0 \right) = N i \left( \langle R | \hat{\rho} | L \rangle - \langle L | \hat{\rho} | R \rangle \right), \]
\[ S_3 \equiv 2 \left( n_3 - n_0 \right) = N \left( \langle R | \hat{\rho} | R \rangle - \langle L | \hat{\rho} | L \rangle \right). \]  

We can now relate the Stokes parameters to the density matrix \( \hat{\rho} \) by the formula
\[ \hat{\rho} = \frac{1}{2} \sum_{i=0}^{3} \frac{S_i}{S_0} \hat{\sigma}_i, \]  

where \( \hat{\sigma}_0 = |R\rangle \langle R| + |L\rangle \langle L| \) is the single qubit identity operator and \( \hat{\sigma}_1 = |R\rangle \langle L| + |L\rangle \langle R|, \hat{\sigma}_2 = i(|L\rangle \langle R| - |R\rangle \langle L|), \) and \( \hat{\sigma}_3 = |R\rangle \langle R| - |L\rangle \langle L| \) are the Pauli spin operators. Thus the measurement of the Stokes parameters can be considered equivalent to a tomographic measurement of the density matrix of an ensemble of single qubits.

2.2. Multiple beam Stokes parameters: Multiple qubit tomography

The generalization of the Stokes scheme to measure the state of multiple photon beams (or multiple qubits) is reasonably straightforward. One should, however, be aware that importance differences exist between the one-photon and the multiple photon cases. Single photons, at least in the
current context, can be described in a purely classical manner, and the density matrix can be related to the purely classical concept of the coherency matrix [28]. For multiple photons one has the possibility of nonclassical correlations occurring, with quintessentially quantum mechanical phenomena such as entanglement being present. We will return to the concept of entanglement and how it may be measured later in this paper.

An $n$-qubit state is characterized by a density matrix which may be written as follows:

$$
\hat{\rho} = \frac{1}{2^n} \sum_{i_1,i_2,\ldots,i_n=0}^{3} r_{i_1,i_2,\ldots,i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \ldots \otimes \hat{\sigma}_{i_n},
$$

(4)

where the $4^n$ parameters $r_{i_1,i_2,\ldots,i_n}$ are real numbers. The normalization property of the density matrices requires that $r_{0,0,\ldots,0} = 1$, and so the density matrix is specified by $4^n - 1$ real parameters. The symbol $\otimes$ represents the tensor product between operators acting on the Hilbert spaces associated with the separate qubits.

As Stokes showed, the state of a single qubit can be determined by taking a set of four projection measurements which are represented by the four operators $\hat{\mu}_0 = |H\rangle\langle H| + |V\rangle\langle V|$, $\hat{\mu}_1 = |H\rangle\langle H|$, $\hat{\mu}_2 = |\bar{D}\rangle\langle \bar{D}|$, $\hat{\mu}_3 = |R\rangle\langle R|$. Similarly, the state of two qubits can be determined by the set of 16 measurements represented by the operators $\hat{\mu}_{i_1} \otimes \hat{\mu}_{j_1}$ ($i,j = 0,1,2,3$). More generally the state of an $n$-qubit system can be determined by $4^n$ measurements given by the operators $\hat{\mu}_{i_1} \otimes \hat{\mu}_{i_2} \otimes \ldots \otimes \hat{\mu}_{i_n}$ ($i_k = 0,1,2,3$ and $k = 1,2,\ldots,n$). This “tree” structure for multi-qubit measurement is illustrated in Fig. 1.

![Fig. 1. Tree diagram representing number and type of measurements necessary for tomography. For a single qubit, the measurements $\{\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3\}$ suffice to reconstruct the state, e.g., measurements of the horizontal, vertical, diagonal and right-circular polarization components, $(H,V,D,R)$. For two qubits, 16 double-coincidence measurements are necessary $(\{\hat{\mu}_0 \hat{\mu}_0, \hat{\mu}_0 \hat{\mu}_1, \ldots, \hat{\mu}_3 \hat{\mu}_3\})$, increasing to 64 three-coincidence measurements for three qubits $(\{\hat{\mu}_0 \hat{\mu}_0 \hat{\mu}_0, \hat{\mu}_0 \hat{\mu}_0 \hat{\mu}_1, \ldots, \hat{\mu}_3 \hat{\mu}_3 \hat{\mu}_3\})$, and so on, as shown.](image)
The proof of this conjecture is reasonably straightforward. The outcome of a measurement is given by the formula
\[
n = \mathcal{N} \text{Tr} \{ \hat{\rho} \hat{\mu} \},
\]
where \( \hat{\rho} \) is the density matrix, \( \hat{\mu} \) is the measurement operator, and \( \mathcal{N} \) is a constant of proportionality which can be determined from the data. Thus in our \( n \)-qubit case the outcomes of the various measurement are
\[
n_{i_1, i_2, \ldots, i_n} = \mathcal{N} \text{Tr} \{ \hat{\rho} (\hat{\mu}_{i_1} \otimes \hat{\mu}_{i_2} \otimes \ldots \otimes \hat{\mu}_{i_n}) \}.
\]
(6)
Substituting from Eq. (4) we obtain
\[
n_{i_1, i_2, \ldots, i_n} = \frac{\mathcal{N}}{2^n} \sum_{j_1, j_2, \ldots, j_n=0}^{3} \text{Tr} \{ \hat{\mu}_{i_1} \hat{\sigma}_{j_1} \} \text{Tr} \{ \hat{\mu}_{i_2} \hat{\sigma}_{j_2} \} \ldots \text{Tr} \{ \hat{\mu}_{i_n} \hat{\sigma}_{j_n} \} r_{i_1, i_2, \ldots, i_n}.
\]
(7)
As can be easily verified, the single qubit measurement operators \( \hat{\mu}_i \) are linear combinations of the Pauli operators \( \hat{\sigma}_j \), i.e., \( \hat{\mu}_i = \sum_{j=0}^{3} \gamma_{ij} \hat{\sigma}_j \), where \( \gamma_{ij} \) are the elements of the matrix
\[
\gamma = \begin{pmatrix}
1 & 0 & 0 & 0 \\
1/2 & 1/2 & 0 & 0 \\
1/2 & 0 & 1/2 & 0 \\
1/2 & 0 & 0 & 1/2
\end{pmatrix}.
\]
(8)
Further, we have the relation \( \text{Tr} \{ \hat{\sigma}_i \hat{\sigma}_j \} = 2 \delta_{ij} \) (where \( \delta_{ij} \) is the Kronecker delta). Hence Eq. (7) becomes
\[
n_{i_1, i_2, \ldots, i_n} = \mathcal{N} \sum_{j_1, j_2, \ldots, j_n=0}^{3} \gamma_{i_1, j_1} \gamma_{i_2, j_2} \ldots \gamma_{i_n, j_n} r_{i_1, i_2, \ldots, i_n}.
\]
(9)
Introducing the left inverse of the matrix \( \gamma \), defined so that \( \sum_{k=0}^{3} (\gamma^{-1})_{ik} \gamma_{kj} = \delta_{ij} \) and whose elements are
\[
\gamma^{-1} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
-1 & 2 & 0 & 0 \\
-1 & 0 & 2 & 0 \\
-1 & 0 & 0 & 2
\end{pmatrix},
\]
(10)
we can find a formula for the parameters \( r_{i_1, i_2, \ldots, i_n} \) in terms of the measured quantities \( n_{i_1, i_2, \ldots, i_n} \), viz.,
\[
\mathcal{N} r_{i_1, i_2, \ldots, i_n} = \sum_{j_1, j_2, \ldots, j_n=0}^{3} (\gamma^{-1})_{i_1, j_1} (\gamma^{-1})_{i_2, j_2} \ldots (\gamma^{-1})_{i_n, j_n} n_{i_1, i_2, \ldots, i_n}
\equiv S_{i_1, i_2, \ldots, i_n}.
\]
(11)
In Eq. (11) we have introduced the $n$-photon Stokes parameter $S_{i_1,i_2,...,i_n}$, defined in an analogous manner to the single photon Stokes parameters given in Eq. (2).

Since, as already noted, $r_{0,0,...,0} = 1$, we can make the identification $S_{0,0,...,0} = N$, and so the density matrix for the $n$-qubit system can be written in terms of the Stokes parameters as follows:

$$\hat{\rho} = \frac{1}{2^n} \sum_{i_1,i_2,...,i_n=0}^{3} S_{i_1,i_2,...,i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \ldots \otimes \hat{\sigma}_{i_n}. \quad (12)$$

This is a recipe for measurement of the density matrices which, assuming perfect experimental conditions and the complete absence of noise, will always work. It is important to realize that the set of four Stokes measurements $\{\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3\}$ are not unique: there may be circumstances in which it is more convenient to use some other set, which is equivalent. A more typical set, at least in optical experiments, is $\hat{\mu}'_0 = |H\rangle\langle H|$, $\hat{\mu}'_1 = |V\rangle\langle V|$, $\hat{\mu}'_2 = |D\rangle\langle D|$, $\hat{\mu}'_3 = |R\rangle\langle R|$.

In the following section we will explore more general schemes for the measurement of two qubits, starting with a discussion, in some detail, of how the measurements are actually performed.

3. Generalized Tomographic Reconstruction of the Polarization State of Two Photons

3.1. Experimental set-up

The experimental arrangement used in our experiments is shown schematically in Fig. 2. An optical system consisting of lasers, polarization elements, and nonlinear optical crystals (collectively characterized for the purposes of this paper as a “black-box,”) is used to generate pairs of qubits in an almost arbitrary quantum state of their polarization degrees of freedom. A full description of this optical system and how such quantum states can be prepared can be found in Ref. [13, 27, 3]*. The output of the black box consists of a pair of beams of light, whose quanta can be measured by means of photodetectors. To project the light beams onto a polarization state of the experimenter’s choosing, three optical elements are placed in the beam

---

*It is important to realize that the entangled photon pairs are produced in a nondeterministic manner: one cannot specify with certainty when a photon pair will be emitted; indeed there is a small probability of generating four or six or higher number of photons. Thus we can only postselectively generate entangled photon pairs: i.e., one only knows that the state was created after if has been measured.
in front of each detector: a polarizer (which transmits only vertically polarized light), a quarter-wave plate, and a half-wave plate. The angles of the fast axes of both of the wave plates can be set arbitrarily, allowing the $|V\rangle$ projection state fixed by the polarizer to be rotated into any polarization state that the experimenter may wish.

![Schematic illustration of the experimental arrangement.](image)

**Fig. 2.** Schematic illustration of the experimental arrangement. QWP stands for quarter-wave plate, HWP for half-wave plate; the angles of both pairs of wave plates can be set independently giving the experimenter four degrees of freedom with which to set the projection state. In the experiment, the polarizers were realized using polarizing prisms, arranged to transmit vertically polarized light.

Using the Jones calculus notation, with the convention,

$$
\begin{pmatrix}
0 \\
1
\end{pmatrix} = |V\rangle, \quad \begin{pmatrix}
1 \\
0
\end{pmatrix} = |H\rangle,
$$

where $|V\rangle$ ($|H\rangle$) is the ket for a vertically (horizontally) polarized beam, the effect of quarter- and half-wave plates whose fast axes are at angles $q$ and $h$ with respect to the vertical axis, respectively, are given by the $2 \times 2$ matrices

$$
\hat{U}_{QWP}(q) = \frac{1}{\sqrt{2}} \begin{pmatrix}
i \cos(2q) & \sin(2q) \\
\sin(2q) & i \cos(2q)
\end{pmatrix}, \quad \hat{U}_{HWP}(q) = \begin{pmatrix}
\cos(2h) & -\sin(2h) \\
-\sin(2h) & \cos(2h)
\end{pmatrix}.
$$

Thus the projection state for the measurement in one of the beams is given by

$$
|\psi^{(1)}_{\text{proj}}(h, q)\rangle = \hat{U}_{QWP}(q) \cdot \hat{U}_{HWP}(h) \cdot \begin{pmatrix}
0 \\
1
\end{pmatrix} = a(h, q)|H\rangle + b(h, q)|V\rangle,
$$
where, neglecting an overall phase, the functions \( a(h, q) \) and \( b(h, q) \) are given by
\[
\begin{align*}
a(h, q) &= \frac{1}{\sqrt{2}} (\sin(2h) - i \sin[2(h - q)]) , \\
b(h, q) &= -\frac{1}{\sqrt{2}} (\cos(2h) + i \cos[2(h - q)]) .
\end{align*}
\]

(14)

The projection state for the two beams is given by
\[
|\psi^{(2)}_{\text{proj}}(h_1, q_1, h_2, q_2)\rangle = |\psi^{(1)}_{\text{proj}}(h_1, q_1)\rangle \otimes |\psi^{(1)}_{\text{proj}}(h_2, q_2)\rangle
\]
\[
= a(h_1, q_1)a(h_2, q_2)|\text{HH}\rangle + a(h_1, q_1)b(h_2, q_2)|\text{HV}\rangle + b(h_1, q_1)a(h_2, q_2)|\text{VH}\rangle + b(h_1, q_1)b(h_2, q_2)|\text{VV}\rangle.
\]

(15)

We shall denote the projection state corresponding to one particular set of wave plate angles \( \{h_{1,\nu}, q_{1,\nu}, h_{2,\nu}, q_{2,\nu}\} \) by the ket \( |\psi_\nu\rangle \);† thus the projection measurement is represented by the operator \( \hat{\mu}_\nu = |\psi_\nu\rangle\langle \psi_\nu| \). Consequently, the average number of coincidence counts that will be observed in a given experimental run is
\[
n_\nu = \mathcal{N} \langle \psi_\nu | \hat{\rho} | \psi_\nu \rangle
\]

(16)

where \( \hat{\rho} \) is the density matrix describing the ensemble of qubits, and \( \mathcal{N} \) is a constant dependent on the photon flux and detector efficiencies. In what follows, it will be convenient to consider the quantities \( s_\nu \) defined by
\[
s_\nu = \langle \psi_\nu | \hat{\rho} | \psi_\nu \rangle .
\]

(17)

### 3.2. Tomographically complete set of measurements

In Section 2 we have given one possible set of projection measurements \( \{|\psi_\nu\rangle \langle \psi_\nu|\} \) which uniquely determine the density matrix \( \hat{\rho} \). However, one can conceive of situations in which these will not be the most convenient set of measurements to make. Here we address the problem of finding other sets of suitable measurements. The smallest number of states required for such measurements can be found by a simple argument: there are 15 real unknown parameters which determine a \( 4 \times 4 \) density matrix, plus there is the single unknown real parameter \( \mathcal{N} \), making a total of 16.

In order to proceed it is helpful to convert the \( 4 \times 4 \) matrix \( \hat{\rho} \) into a 16-dimensional column vector. To do this we use a set of 16 linearly

---

†Here the first subscript on the wave plate angle refers one of the two photon beams; the second subscript distinguishes which of the 16 different experimental states is under consideration.
independent $4 \times 4$ matrices $\{\hat{\Gamma}_\nu\}$ which have the following mathematical properties:

$$\text{Tr} \left\{ \hat{\Gamma}_\nu \cdot \hat{\Gamma}_\mu \right\} = \delta_{\nu,\mu}, \quad \hat{A} = \sum_{\nu=1}^{16} \hat{\Gamma}_\nu \text{Tr} \left\{ \hat{\Gamma}_\nu \cdot \hat{A} \right\} \quad \forall \hat{A}, \quad (18)$$

where $\hat{A}$ is an arbitrary $4 \times 4$ matrix. Finding a set of $\hat{\Gamma}_\nu$ matrices is in fact reasonably straightforward: for example, the set of (appropriately normalized) generators of the Lie algebra $\text{SU}(2) \otimes \text{SU}(2)$ fulfill the required criteria (for reference, we list this set in Appendix A). These matrices are of course simply a relabeling of the two-qubit Pauli matrices $\hat{\sigma}_i \otimes \hat{\sigma}_j \ (i, j = 0, 1, 2, 3)$ discussed above. Using these matrices the density matrix can be written as

$$\hat{\rho} = \sum_{\nu=1}^{16} \hat{\Gamma}_\nu r_\nu, \quad (19)$$

where $r_\nu$ is the $\nu$th element of a 16-element column vector, given by the formula

$$r_\nu = \text{Tr} \left\{ \hat{\Gamma}_\nu \cdot \hat{\rho} \right\} \quad (20)$$

Substituting Eq. (19) into Eq. (16), we obtain the following linear relationship between the measured coincidence counts $n_\nu$ and the elements of the vector $r_\mu$:

$$n_\nu = N \sum_{\mu=1}^{16} B_{\nu,\mu} r_\mu, \quad (21)$$

where the $16 \times 16$ matrix $B_{\nu,\mu}$ is given by

$$B_{\nu,\mu} = \langle \psi_\nu | \hat{\Gamma}_\mu | \psi_\nu \rangle. \quad (22)$$

Immediately we find a necessary and sufficient condition for the completeness of the set of tomographic states $\{ |\psi_\nu \rangle \}$: if the matrix $B_{\nu,\mu}$ is nonsingular, then Eq. (21) can be inverted to give

$$r_\nu = (N)^{-1} \sum_{\mu=1}^{16} (B^{-1})_{\nu,\mu} n_\mu. \quad (23)$$

The set of 16 tomographic states that we employed are given in Table 1. They can be shown to satisfy the condition that $B_{\nu,\mu}$ is nonsingular. By no means are these states unique in this regard: these were the states chosen principally for experimental convenience.
Table 1: The tomographic analysis states used in our experiments. The number of coincidence counts measured in projections measurements provide a set of 16 data that allow the density matrix of the state of the two modes to be estimated. We have used the notation $|D\rangle \equiv (|H\rangle + |V\rangle) / \sqrt{2}$, $|L\rangle \equiv (|H\rangle + i|V\rangle) / \sqrt{2}$ and $|R\rangle \equiv (|H\rangle - i|V\rangle) / \sqrt{2}$. Note that, when the measurement are taken in the order given by the table, only one wave plate angle has to be changed between measurements.

These states can be realized by setting specific values of the half- and quarter-wave plate angles. The appropriate values of these angles (measured from the vertical) are given in Table 1. Note that overall phase factors do not affect the results of projection measurements.

Substituting Eq. (23) into Eq. (19), we find that

$$\hat{\rho} = (\mathcal{N})^{-1} \sum_{\nu=1}^{16} \hat{M}_{\nu} n_{\nu} = \sum_{\nu=1}^{16} \hat{M}_{\nu} s_{\nu},$$

(24)

where the sixteen $4 \times 4$ matrices $\hat{M}_{\nu}$ are defined by

$$\hat{M}_{\nu} = \sum_{\nu=1}^{16} (B^{-1})_{\nu,\mu} \hat{\Gamma}_{\mu}.$$ 

(25)

The introduction of the $\hat{M}_{\nu}$ matrices allows a compact form of linear tomographic reconstruction, Eq. (24), which will be most useful in the error
analysis that follows. These $\hat{M}_\nu$ matrices, valid for our set of tomographic states, are listed in Appendix B, together with some of their important properties. We can use one of these properties, Eq. (80), to obtain the value of the unknown quantity $\mathcal{N}$. That relationship implies

$$\sum_\nu \text{Tr}\{\hat{M}_\nu\} |\psi_\nu\rangle\langle \psi_\nu| = \hat{\rho}.$$  \hspace{1cm} (26)$$

Taking the trace of this formula, and multiplying by $\mathcal{N}$ we obtain:

$$\sum_\nu \text{Tr}\{\hat{M}_\nu\} n_\nu = \mathcal{N}.\hspace{1cm} (27)$$

For our set of tomographic states, it can be shown that

$$\text{Tr}\{\hat{M}_\nu\} = \begin{cases} 1 & \text{if } \nu = 1, 2, 3, 4 \\ 0 & \text{if } \nu = 5, \ldots, 16; \end{cases} \hspace{1cm} (28)$$

hence the value of the unknown parameter $\mathcal{N}$ in our experiments is given by:

$$\mathcal{N} = \sum_{\nu=1}^{4} n_\nu = \mathcal{N} (\langle HH | \hat{\rho} | HH \rangle + \langle HV | \hat{\rho} | HV \rangle + \langle VH | \hat{\rho} | VH \rangle + \langle VV | \hat{\rho} | VV \rangle).\hspace{1cm} (29)$$

Thus we obtain the final formula for the tomographic reconstruction of the density matrices of our states:

$$\hat{\rho} = \left( \sum_{\nu=1}^{16} \hat{M}_\nu n_\nu \right) / \left( \sum_{\nu=1}^{4} n_\nu \right). \hspace{1cm} (30)$$

As an example, the following set of 16 counts were taken for the purpose of tomographically determining the density matrix for an ensemble of qubits all prepared in a specific quantum state: $n_1 = 34749$, $n_2 = 324$, $n_3 = 35805$, $n_4 = 444$, $n_5 = 16324$, $n_6 = 17521$, $n_7 = 13441$, $n_8 = 16901$, $n_9 = 17932$, $n_{10} = 32028$, $n_{11} = 15132$, $n_{12} = 17238$, $n_{13} = 13171$, $n_{14} = 17170$, $n_{15} = 16722$, $n_{16} = 33586$. Applying Eq. (29) we find

$$\hat{\rho} = \begin{pmatrix}
0.4872 & -0.0042 + i0.0114 & -0.0098 - i0.0178 & 0.5192 + i0.0380 \\
-0.0042 - i0.0114 & 0.0045 & 0.0271 - i0.0146 & -0.0648 - i0.0076 \\
-0.0098 + i0.0178 & 0.0271 + i0.0146 & 0.0062 & -0.0695 + i0.0134 \\
0.5192 - i0.0380 & -0.0648 + i0.0076 & -0.0695 - i0.0134 & 0.5020
\end{pmatrix}. \hspace{1cm} (30)$$

This matrix is shown graphically in Fig. 3(left).

Note that, by construction, the density matrix is normalized, i.e., $\text{Tr}\{\hat{\rho}\} = 1$ and Hermitian, i.e., $\hat{\rho}^\dagger = \hat{\rho}$. However, when one calculates the
eigenvalues of this measured density matrix, one finds the values 1.02155, 0.0681238, -0.065274, and -0.024396; and also, $\text{Tr}\{\hat{\rho}^2\} = 1.053$. Density matrices for all physical states must have the property of positive semidefiniteness, which (in conjunction with the normalization and Hermiticity properties) implies that all of the eigenvalues must lie in the interval $[0, 1]$, their sum being 1; this in turn implies that $0 \leq \text{Tr}\{\hat{\rho}^2\} \leq 1$. Clearly, the density matrix reconstructed above by linear tomography violates these condition. From our experience of tomographic measurements of various mixed and entangled states prepared experimentally, this seems to happen roughly 75% of the time for low entropy, highly entangled states; it seems to have a higher probability of producing the correct result for states of higher entropy, but the cautious experimenter should check every time. The obvious culprit for this problem is experimental inaccuracies and statistical fluctuations of coincidence counts, which mean that the actual numbers of counts recorded in a real experiment differ from those that can be calculated by Eq. (16). Thus the linear reconstruction is of limited value for states of low entropy (which are of most experimental interest because of their application to quantum information technology); however, as we shall see, the linear approach does provide a useful starting point for the numerical optimization approach to density matrix estimation which we will discuss in the next section.

4. Maximum Likelihood Estimation

As mentioned in Section 3, the tomographic measurement of density matrices can produce results which violate important basic properties such as positivity. To avoid this problem, the maximum likelihood estimation of density matrices may be employed. Here we describe a simple realization of this technique.

4.1. Basic approach

Our approach to the maximum likelihood estimation of the density matrix is as follows.

(i) Generate a formula for an explicitly “physical” density matrix, i.e., a matrix that has the three important properties of normalization, Hermiticity, and positivity. This matrix will be a function of 16 real variables (denoted $\{t_1, t_2, \ldots, t_{16}\}$). We will denote the matrix as $\hat{\rho}_p(t_1, t_2, \ldots, t_{16})$.

(ii) Introduce a “likelihood function” which quantifies how good the density matrix $\hat{\rho}_p(t_1, t_2, \ldots, t_{16})$ is in relation to the experimental data.
This likelihood function is a function of the 16 real parameters $t_\nu$ and of the 16 experimental data $n_\nu$. We will denote this function as $\mathcal{L}(t_1, t_2, \ldots, t_{16}; n_1, n_2, \ldots, n_{16})$.

(iii) Using standard numerical optimization techniques, find the optimum set of variables $\{t_1^{(\text{opt})}, t_2^{(\text{opt})}, \ldots, t_{16}^{(\text{opt})}\}$ for which the function $\mathcal{L}(t_1, t_2, \ldots, t_{16}; n_1, n_2, \ldots, n_{16})$ has its maximum value. The best estimate for the density matrix is then $\hat{\rho}(t_1^{(\text{opt})}, t_2^{(\text{opt})}, \ldots, t_{16}^{(\text{opt})})$.

The details of how these three steps can be carried out are described in the next three subsections.

4.2. Physical density matrices

The property of non-negative definiteness for any matrix $\hat{G}$ is written mathematically as

$$\langle \psi | \hat{G} | \psi \rangle \geq 0 \quad \forall |\psi\rangle.$$  \hspace{1cm} (31)

Any matrix that can be written in the form $\hat{G} = \hat{T}^\dagger \hat{T}$ must be non-negative definite. To see that this is the case, substitute into Eq. (31)

$$\langle \psi | \hat{T}^\dagger \hat{T} | \psi \rangle = \langle \psi' | \psi' \rangle \geq 0,$$  \hspace{1cm} (32)

where we have defined $|\psi'\rangle = \hat{T} |\psi\rangle$. Furthermore, $(\hat{T}^\dagger \hat{T})^\dagger = \hat{T}^\dagger (\hat{T}^\dagger)^\dagger = \hat{T}^\dagger \hat{T}$, i.e., $\hat{G} = \hat{T}^\dagger \hat{T}$ must be Hermitian. To ensure normalization, one can simply divide by the trace: thus the matrix $\hat{g}$ given by the formula

$$\hat{g} = \hat{T}^\dagger \hat{T} / \text{Tr}\{\hat{T}^\dagger \hat{T}\}$$  \hspace{1cm} (33)

has all three of the mathematical properties that we require for density matrices.

For the two-qubit system, we have a $4 \times 4$ density matrix with 15 independent real parameters. Since it will be useful to be able to invert relation (33), it is convenient to choose a tridiagonal form for $\hat{T}$:

$$\hat{T}(t) = \begin{pmatrix}
    t_1 & 0 & 0 & 0 \\
    t_5 + it_6 & t_2 & 0 & 0 \\
    t_{11} + it_{12} & t_7 + it_8 & t_3 & 0 \\
    t_{15} + it_{16} & t_{13} + it_{14} & t_9 + it_{10} & t_4
\end{pmatrix}. \hspace{1cm} (34)$$

Thus the explicitly “physical” density matrix $\hat{\rho}_p$ is given by the formula

$$\hat{\rho}_p(t) = \hat{T}^\dagger(t) \hat{T}(t) / \text{Tr}\{\hat{T}^\dagger(t) \hat{T}(t)\}.$$  \hspace{1cm} (35)
For future reference, the inverse relationship, by which the elements of \(\hat{T}\) can be expressed in terms of the elements of \(\hat{\rho}\), is as follows:

\[
\hat{T} = \begin{pmatrix}
\sqrt{\frac{\Delta}{\mathcal{M}_{11}^{(1)}}} & 0 & 0 & 0 \\
\frac{\mathcal{M}_{13}^{(1)}}{\sqrt{\mathcal{M}_{11}^{(1)}\mathcal{M}_{12}^{(2)}}} & \sqrt{\frac{\mathcal{M}_{11}^{(1)}}{\mathcal{M}_{11,22}^{(2)}}} & 0 & 0 \\
\frac{\mathcal{M}_{12,23}^{(2)}}{\sqrt{\rho_{14}\mathcal{M}_{11,22}^{(2)}}} & \frac{\mathcal{M}_{11,23}^{(2)}}{\sqrt{\rho_{14}\mathcal{M}_{11,22}^{(2)}}} & \sqrt{\frac{\mathcal{M}_{11,22}^{(2)}}{\rho_{44}}} & 0 \\
\frac{\rho_{41}}{\sqrt{\rho_{14}}} & \frac{\rho_{42}}{\sqrt{\rho_{14}}} & \frac{\rho_{43}}{\sqrt{\rho_{14}}} & \sqrt{\rho_{44}}
\end{pmatrix}.
\]

(36)

Here we have used the notation \(\Delta = \text{Det}(\hat{\rho})\); \(\mathcal{M}_{ij}^{(1)}\) is the first minor of \(\hat{\rho}\), i.e., the determinant of the \(3 \times 3\) matrix formed by deleting the \(i\)-th row and \(j\)-th column of \(\hat{\rho}\); \(\mathcal{M}_{ij,kl}^{(2)}\) is the second minor of \(\hat{\rho}\), i.e., the determinant of the \(2 \times 2\) matrix formed by deleting the \(i\)-th and \(k\)-th rows and \(j\)-th and \(l\)-th columns of \(\hat{\rho}\) \((i \neq k \text{ and } j \neq l)\).

### 4.3. The likelihood function

The measurement data consists of a set of 16 coincidence counts \(n_\nu\) \((\nu = 1, 2, \ldots, 16)\) whose expected value is \(\bar{n}_\nu = \mathcal{N}(\psi_\nu | \hat{\rho} | \psi_\nu)\). Let us assume that the noise on these coincidence measurements has a Gaussian probability distribution. Thus the probability of obtaining a set of 16 counts \(\{n_1, n_2, \ldots, n_{16}\}\) is

\[
P(n_1, n_2, \ldots, n_{16}) = \frac{1}{\mathcal{N}_{\text{norm}}} \prod_{\nu=1}^{16} \exp\left[-\frac{(n_\nu - \bar{n}_\nu)^2}{2\sigma_\nu^2}\right],
\]

(37)

where \(\sigma_\nu\) is the standard deviation for \(\nu\)-th coincidence measurement (given approximately by \(\sqrt{n_\nu}\)) and \(\mathcal{N}_{\text{norm}}\) is the normalization constant. For our physical density matrix \(\hat{\rho}_p\) the number of counts expected for the \(\nu\)-th measurement is

\[
\bar{n}_\nu(t_1, t_2, \ldots, t_{16}) = \mathcal{N}(\psi_\nu | \hat{\rho}_p(t_1, t_2, \ldots, t_{16}) | \psi_\nu).
\]

(38)

Thus the likelihood that the matrix \(\hat{\rho}_p(t_1, t_2, \ldots, t_{16})\) could produce the measured data \(\{n_1, n_2, \ldots, n_{16}\}\) is

\[
P(n_1, n_2, \ldots, n_{16}) = \frac{1}{\mathcal{N}_{\text{norm}}} \prod_{\nu=1}^{16} \exp\left[-\frac{(\mathcal{N}(\psi_\nu | \hat{\rho}_p(t_1, t_2, \ldots, t_{16}) | \psi_\nu) - n_\nu)^2}{2\mathcal{N}(\psi_\nu | \hat{\rho}_p(t_1, t_2, \ldots, t_{16}) | \psi_\nu))}\right],
\]

(39)
where $\mathcal{N} = \sum_{\nu=1}^{4} N_\nu$.

Rather than find maximum value of $P(t_1, t_2, \ldots, t_{16})$ it simplifies things somewhat to find the maximum of its logarithm (which is mathematically equivalent).‡ Thus the optimization problem reduces to finding the minimum of the following function:

$$
L(t_1, t_2, \ldots, t_{16}) = \sum_{\nu=1}^{16} \frac{(N\langle \psi_\nu | \hat{\rho}_p (t_1, t_2, \ldots, t_{16}) | \psi_\nu \rangle - n_\nu)^2}{2N\langle \psi_\nu | \hat{\rho}_p (t_1, t_2, \ldots, t_{16}) | \psi_\nu \rangle}.
$$

(40)

This is the “likelihood” function that we employed in our numerical optimization routine.

4.4. Numerical optimization

We used the Mathematica 4.0 routine FindMinimum which executes a multidimensional Powell direction set algorithm (see Ref. [18] for a description of this algorithm). To execute this routine, one requires an initial estimate for the values of $t_1, t_2, \ldots, t_{16}$. For this, we used the tomographic estimate of the density matrix in the inverse relation (36), allowing us to determine a set of values for $t_1, t_2, \ldots, t_{16}$. Since the tomographic density matrix may not be nonnegative definite, the values of the $t_\nu$’s deduced in this manner are not necessarily real. Thus for our initial guess we used the real parts of the $t_\nu$’s deduced from the tomographic density matrix.

For the example given in Section 2, the maximum likelihood estimate is

$$
\hat{\rho} = \begin{pmatrix}
0.5069 & -0.0239 + i0.0106 & -0.0412 - i0.0221 & 0.4833 + i0.0329 \\
-0.0239 - i0.0106 & 0.0048 & 0.0023 + i0.0019 & -0.0296 - i0.0077 \\
-0.0412 + i0.0221 & 0.0023 - i0.0019 & 0.0045 & -0.0425 + i0.0192 \\
0.4833 - i0.0329 & -0.0296 + i0.0077 & -0.0425 - i0.0192 & 0.4839
\end{pmatrix}.
$$

(41)

This matrix is illustrated in Fig. 3(right). In this case, the matrix has eigenvalues 0.986022, 0.0139777, 0, and 0; and $\text{Tr}\{\hat{\rho}^2\} = 0.972435$, indicating that, while the linear reconstruction gave a nonphysical density matrix, the maximum likelihood reconstruction gives a legitimate density matrix.

5. Error Analysis

In this section we present an analysis of the errors inherent in the tomographic scheme described in Section 3. Two sources of error are found to

‡Note that here we neglect the dependence of the normalization constant on $t_1, t_2, \ldots, t_{16}$, which only weakly affects solution for the most likely state.
be important: the shot noise error in the measured coincidence counts $n_\nu$ and the uncertainty in the settings of the angles of the wave plates used to make the tomographic projection states. We will analyze these two sources separately.

In addition to determining the density matrix of a pair of qubits, one is often also interested in quantities derived from the density matrix, such as the entropy or the entanglement of formation. For completeness, we will also derive the errors in some of these quantities.

5.1. Errors due to count statistics

From Eq. (29) we see that the density matrix is specified by a set of 16 parameters $s_\nu$ defined by

$$s_\nu = n_\nu / N,$$

where $n_\nu$ are the measured coincidence counts and $N = \sum_{\nu=1}^{4} n_\nu$. We can determine the errors in $s_\nu$ using the following formula [16]:

$$\overline{\delta s_\nu \delta s_\mu} = \sum_{\lambda,\kappa=1}^{16} \left( \frac{\partial s_\nu}{\partial n_\lambda} \right) \left( \frac{\partial s_\mu}{\partial n_\kappa} \right) \delta n_\lambda \delta n_\kappa,$$
where the overbar denotes the ensemble average of the random uncertainties \( \delta s_\nu \) and \( \delta n_\lambda \). The measured coincidence counts \( n_\lambda \) are statistically independent Poissonian random variables, which implies the following relation:

\[
\overline{\delta n_\lambda \delta n_\kappa} = n_\lambda \delta_{\lambda,\kappa}, \tag{44}
\]

where \( \delta_{\lambda,\kappa} \) is the Kronecker delta.

Taking the derivative of Eq. (42), we find that

\[
\frac{\partial s_\mu}{\partial n_\nu} = \frac{1}{N^2} \delta_{\mu \nu} - \frac{n_\mu}{N^2} D_\nu, \tag{45}
\]

where

\[
D_\nu = \sum_{\lambda=1}^{4} \delta_{\lambda,\nu} = \begin{cases} 
1 & \text{if } 1 \leq \nu \leq 4 \\
0 & \text{if } 5 \leq \nu \leq 16. 
\end{cases} \tag{46}
\]

Substituting Eq. (45) into Eq. (43) and using Eq. (44), we obtain the result

\[
\overline{\delta s_\nu \delta s_\mu} = \frac{n_\mu}{N^2} \delta_{\nu,\mu} + \frac{n_\nu n_\mu}{N^3} (1 - D_\mu - D_\nu). \tag{47}
\]

In most experimental circumstances \( Ng1 \), and so the second term in Eq. (47) is negligibly small in comparison to the first. We shall therefore ignore it, and use the approximate expression in the subsequent discussion:

\[
\overline{\delta s_\nu \delta s_\mu} \approx \frac{n_\mu}{N^2} \delta_{\nu,\mu} \equiv \frac{s_\mu}{N} \delta_{\nu,\mu}. \tag{48}
\]

5.2. Errors due to angular settings uncertainties

Using the formula (17) for the parameters \( s_\nu \) we can find the dependence of the measured density matrix on errors in the tomographic states. The derivative of \( s_\nu \) with respect to some generic wave plate settings angle \( \theta \) is

\[
\frac{\partial s_\nu}{\partial \theta} = \left\{ \frac{\partial}{\partial \theta} \langle \psi_\nu | \hat{\rho} | \psi_\nu \rangle \right\} + \langle \psi_\nu | \hat{\rho} \left\{ \frac{\partial}{\partial \theta} | \psi_\nu \rangle \right\}, \tag{49}
\]

where \( | \psi_\nu \rangle \) is the ket of the \( \nu \)th projection state [see Eq. (15)]. Substituting Eq. (24) we find

\[
\frac{\partial s_\nu}{\partial \theta} = \sum_{\mu=1}^{16} s_\mu \left\{ \frac{\partial}{\partial \theta} \langle \psi_\nu | \hat{\rho} | \psi_\nu \rangle \right\} + \langle \psi_\nu | \hat{M}_\mu | \psi_\nu \rangle \left\{ \frac{\partial}{\partial \theta} | \psi_\nu \rangle \right\}. \tag{50}
\]

For convenience, we shall label the four wave plate angles \( \{h_1, q_1, h_2, q_2\} \), which specify the \( \nu \)th state by \( \{\theta_{\nu,1}, \theta_{\nu,2}, \theta_{\nu,3}, \theta_{\nu,4}\} \), respectively. Clearly the \( \mu \)th state does not depend on any of the \( \nu \)th set
of angles. Thus we obtain the following expression for the derivatives of \( s_\nu \) with respect to wave plate settings:

\[
\frac{\partial s_\nu}{\partial \theta_{\lambda,i}} = \delta_{\nu,\lambda} \sum_{\mu=1}^{16} s_\mu f^{(i)}_{\nu,\mu},
\]

(51)

where

\[
f^{(i)}_{\nu,\mu} = \left\{ \frac{\partial}{\partial \theta_{\nu,i}} \langle \psi_\nu | \hat{M}_\mu | \psi_\nu \rangle + \langle \psi_\nu | \hat{M}_\mu \left\{ \frac{\partial}{\partial \theta_{\nu,i}} | \psi_\nu \rangle \right\} \right\}.
\]

(52)

The 1024 quantities \( f^{(i)}_{\nu,\mu} \) can be determined by taking the derivatives of the functional forms of the tomographic states given by Eqs. (14) and (15), and evaluating those derivatives at the appropriate values of the arguments (see Table 1).

The errors in the angles are assumed to be uncorrelated, as would be the case if each wave plate were adjusted for each of the 16 measurements. In reality, for qubit experiments, only one or two of the four wave plates are adjusted between one measurement and the next. However, the assumption of uncorrelated angular errors greatly simplifies the calculation (which is, after all, only an estimate of the errors), and seems to produce reasonable figures for our error bars.\(^8\) Thus, with the assumption

\[
\overline{\delta \theta_{\nu,i} \delta \theta_{\mu,j}} = \delta_{\nu,\mu} \delta_{i,j} (\Delta \theta)^2
\]

(53)

(where \( \Delta \theta \) is the rms uncertainty in the setting of the wave plate, with an estimated value of 0.25° for our apparatus), we obtain the following expression for the errors in \( s_\nu \) due to angular settings:

\[
\overline{\delta s_\nu \delta s_\mu} = \delta_{\nu,\mu} \sum_{i=1}^{4} \sum_{\varepsilon,\lambda=1}^{16} f^{(i)}_{\nu,\varepsilon} f^{(i)}_{\nu,\lambda} s_\varepsilon s_\lambda.
\]

(54)

Combining Eqs. (54) and (48) we obtain the following formula for the total error in the quantities \( s_\nu \):

\[
\overline{\delta s_\nu \delta s_\mu} = \delta_{\nu,\mu} \Lambda_\nu
\]

(55)

where

\[
\Lambda_\nu = \left[ \frac{s_\nu}{N} + \sum_{i=1}^{4} \sum_{\varepsilon,\lambda=1}^{16} f^{(i)}_{\nu,\varepsilon} f^{(i)}_{\nu,\lambda} s_\varepsilon s_\lambda \right].
\]

(56)

\(^8\)In other experimental circumstances, such as the measurement of the joint state of two spin-1/2 particles, the tomography would be realized by performing unitary operations on the spins prior to measurement. In this case, an assumption analogous to ours will be wholly justified.
These 16 quantities can be calculated using the parameters $s_\nu$ and the constants $f_{\nu,\varepsilon}^{(i)}$. Note that the same result can be obtained by assuming a priori that the errors in $s_\nu$ are all uncorrelated, with $\Lambda_\nu = \delta s_\nu^2$; the more rigorous treatment given here is necessary, however, to demonstrate this fact. For a typical number of counts, say $N = 10000$, it is found that the contribution of errors from the two causes is roughly comparable; for larger numbers of counts, the angular settings will become the dominant source of error.

Based on these results, the errors in the values of the various elements of the density matrix estimated by the linear tomographic technique described in Section 3 are as follows:

$$
(\Delta \rho_{i,j})^2 = \sum_{\nu,\mu=1}^{16} \frac{\partial \rho_{i,j}}{\partial s_\nu} \frac{\partial \rho_{i,j}}{\partial s_\mu} \delta s_\nu \delta s_\mu = \sum_{\nu=1}^{16} (M_{\nu(i,j)})^2 \Lambda_\nu
$$

(57)

where $M_{\nu(i,j)}$ is the $i,j$ element of the matrix $\hat{M}_\nu$.

A convenient way in which to estimate errors for a maximum likelihood tomographic technique (rather than a linear tomographic technique) is to employ the above formulas, with the slight modification that the parameter $s_\nu$ should be recalculated from Eq. (17) using the estimated density matrix $\hat{\rho}_{\text{est}}$. This does not take into account errors inherent in the maximum likelihood technique itself.

### 5.3. Errors in quantities derived from the density matrix

When calculating the propagation of errors, it is actually more convenient to use the errors in the $s_\nu$ parameters [given by Eq. (56)], rather than the errors in the elements of density matrix itself (which have non-negligible correlations).

#### 5.3.1. von Neumann entropy

The von Neumann entropy is an important measure of the purity of a quantum state $\hat{\rho}$. It is defined by [17]

$$
S = -\text{Tr} \{ \hat{\rho} \log_2 (\hat{\rho}) \} = -\sum_{a=1}^{4} p_a \log_2 p_a,
$$

(58)

where $p_a$ is an eigenvalue of $\hat{\rho}$, i.e.,

$$
\hat{\rho} |\phi_a\rangle = p_a |\phi_a\rangle,
$$

(59)
|φ_a⟩ being the a\text{th} eigenstate (a = 1, \ldots, 4). The error in this quantity is given by
\[
(ΔS)^2 = \sum_{\nu=1}^{16} \left( \frac{\partial S}{\partial s_\nu} \right)^2 Λ_\nu. \tag{60}
\]

Applying the chain rule, we find
\[
\left( \frac{\partial S}{\partial s_\nu} \right) = \sum_{a=1}^{4} \left( \frac{\partial p_a}{\partial s_\nu} \right) \left( \frac{\partial S}{\partial p_a} \right). \tag{61}
\]

The partial differential of an eigenvalue can be easily found by perturbation theory. As is well known (e.g., [22]) the change in the eigenvalue \( \lambda_a \) of a matrix \( \hat{W} \) due to a perturbation in the matrix \( \delta \hat{W} \) is
\[
\delta \lambda_a = \langle \phi_a | \delta \hat{W} | \phi_a \rangle, \tag{62}
\]
where \( |\phi_a⟩ \) is the eigenvector of \( \hat{W} \) corresponding to the eigenvalue \( \lambda_a \). Thus the derivative of \( \lambda_a \) with respect to some variable \( x \) is given by
\[
\frac{\partial \lambda_a}{\partial x} = \langle \phi_a | \frac{\partial \hat{W}}{\partial x} | \phi_a \rangle. \tag{63}
\]

Since \( \hat{\rho} = \sum_{\nu=1}^{16} \hat{M}_\nu s_\nu \), we find that
\[
\frac{\partial p_a}{\partial s_\nu} = \langle \phi_a | \hat{M}_\nu | \phi_a \rangle \tag{64}
\]
and so, taking the derivative of Eq. (58), Eq. (61) becomes
\[
\left( \frac{\partial S}{\partial s_\nu} \right) = -\sum_{a=1}^{4} \langle \phi_a | \hat{M}_\nu | \phi_a \rangle \left[ 1 + \ln p_a \right] \frac{1}{\ln 2}. \tag{65}
\]
Hence
\[
(ΔS)^2 = \sum_{\nu=1}^{16} \left( \sum_{a=1}^{4} \langle \phi_a | \hat{M}_\nu | \phi_a \rangle \left[ 1 + \ln p_a \right] \frac{1}{\ln 2} \right)^2 Λ_\nu. \tag{66}
\]

For the experimental example given above, \( S = 0.106 \pm 0.049 \).

5.3.2. Linear entropy

The “linear entropy” is used to quantify the degree of mixture of a quantum state in an analytically convenient form, although unlike the von Neumann entropy it has no direct information theoretic implications. In a normalized
form (defined so that its value lies between zero and one), the linear entropy for a two-qubit system is defined by

\[
P = \frac{4}{3} \left(1 - \text{Tr} \{ \hat{\rho}^2 \} \right) = \frac{4}{3} \left(1 - \sum_{a=1}^{4} p_a^2 \right).
\]  

(67)

To calculate the error in this quantity, we need the following partial derivative:

\[
\frac{\partial P}{\partial s_\nu} = -\frac{8}{3} \sum_{a=1}^{4} p_a \frac{\partial p_a}{\partial s_\nu} = -\frac{8}{3} \sum_{a=1}^{4} p_a \langle \phi_a | \hat{M}_\nu | \phi_a \rangle
\]

\[
= -\frac{8}{3} \text{Tr} \{ \hat{\rho} \hat{M}_\nu \} = -\frac{8}{3} \sum_{\mu=1}^{16} \text{Tr} \left\{ \hat{M}_\mu \hat{M}_\nu \right\} s_\mu
\]

(68)

Hence the error in the linear entropy is

\[
(\Delta P)^2 = \sum_{\nu=1}^{16} \left( \frac{\partial P}{\partial s_\nu} \right)^2 \Lambda_\nu = \sum_{\nu=1}^{16} \left( \frac{8}{3} \sum_{\mu=1}^{16} \text{Tr} \left\{ \hat{M}_\mu \hat{M}_\nu \right\} s_\mu \right)^2 \Lambda_\nu.
\]  

(69)

For the example given in Sections 3 and 4, \( P = 0.037 \pm 0.026 \).

5.3.3. Concurrence, entanglement of formation and tangle

The concurrence, entanglement of formation, and tangle are measures of the quantum coherence properties of a mixed quantum state [29]. For two qubits,\(^*\) concurrence is defined as follows: consider the non-Hermitian matrix \( \hat{R} = \hat{\rho} \hat{\Sigma} \hat{\rho}^T \hat{\Sigma} \) where the superscript \( T \) denotes transpose and the “spin flip matrix” \( \hat{\Sigma} \) is defined by

\[
\hat{\Sigma} = \begin{pmatrix}
0 & 0 & 0 & -1 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 0
\end{pmatrix}.
\]  

(70)

Note that the definition of \( \hat{\Sigma} \) depends on the basis chosen; we have assumed here the “computational basis” \( \{|HH\rangle, |HV\rangle, |VH\rangle, |VV\rangle\} \). In what

\(^*\)The analysis in this subsection applies to the two-qubit case only. Measures of entanglement for mixed \( n \)-qubit systems are a subject of on-going research: see, for example, [25] for a recent survey. It may be possible to measure entanglement directly, without quantum state tomography; this possibility was investigated in [21].
On the Measurement of Qubits

follows, it will be convenient to write $\hat{R}$ in the following form:

$$\hat{R} = \frac{1}{2} \sum_{\mu,\nu=1}^{16} \hat{q}_{\mu,\nu} s_{\mu} s_{\nu}, \quad (71)$$

where $\hat{q}_{\mu,\nu} = M_{\mu} \hat{\Sigma} M_{\nu}^T \hat{\Sigma} + M_{\nu} \hat{\Sigma} M_{\mu}^T \hat{\Sigma}$. The left and right eigenstates and eigenvalues of the matrix $\hat{R}$ we shall denote by $\langle \xi_a | \ , | \zeta_a \rangle$, and $r_a$, respectively, i.e.,

$$\langle \xi_a | \hat{R} = r_a \langle \xi_a | , \hat{R} | \zeta_a \rangle = r_a | \zeta_a \rangle. \quad (72)$$

We shall assume that these eigenstates are normalized in the usual manner for biorthogonal expansions, i.e., $\langle \xi_a | \zeta_b \rangle = \delta_{a,b}$. Further we shall assume that the eigenvalues are numbered in decreasing order, so that $r_1 \geq r_2 \geq r_3 \geq r_4$. The concurrence is then defined by the formula

$$C = \max \left\{ 0, \sqrt{r_1} - \sqrt{r_2} - \sqrt{r_3} - \sqrt{r_4} \right\} = \max \left\{ 0, \sum_{a=1}^{4} \text{sgn} \left( \frac{3}{2} - a \right) \sqrt{r_a} \right\},$$

where $\text{sgn}(x) = 1$ if $x > 0$ and $\text{sgn}(x) = -1$ if $x < 0$. The tangle is given by $T = C^2$ and the entanglement of formation by

$$E = h \left( \frac{1 + \sqrt{1 - C^2}}{2} \right), \quad (73)$$

where $h(x) = -x \log_2 x - (1 - x) \log_2 (1 - x)$. Because $h(x)$ is a monotonically increasing function, these three quantities are to some extent equivalent measures of the entanglement of a mixed state.

To calculate the errors in these rather complicated functions, we must employ the perturbation theory for non-Hermitian matrices (see Appendix C for more details). We need to evaluate the following partial derivative:

$$\frac{\partial C}{\partial s_{\nu}} = \sum_{a=1}^{4} \text{sgn} \left( \frac{3}{2} - a \right) \frac{1}{2\sqrt{r_a}} \frac{\partial r_a}{\partial s_{\nu}} = \sum_{a=1}^{4} \text{sgn} \left( \frac{3}{2} - a \right) \frac{1}{2\sqrt{r_a}} \langle \xi_a | \hat{R} \frac{\partial \hat{R}}{\partial s_{\nu}} | \zeta_a \rangle$$

$$= \sum_{a=1}^{4} \sum_{\mu=1}^{16} \text{sgn} \left( \frac{3}{2} - a \right) \frac{1}{2\sqrt{r_a}} \langle \xi_a | \hat{q}_{\mu,\nu} s_{\mu} | \zeta_a \rangle, \quad (74)$$

where the function $\text{sgn}(x)$ is the sign of the quantity $x$: it takes the value 1 if $x > 0$ and $-1$ if $x < 0$. Thus $\text{sgn}(3/2 - a)$ is equal to $+1$ if $a = 1$ and
−1 if \( a = 2, 3, \) or 4. Hence the error in the concurrence is

\[
(\Delta C)^2 = \sum_{\nu=1}^{16} \left( \frac{\partial C}{\partial s_{\nu}} \right)^2 \Lambda_{\nu}
\]

\[
= \sum_{\nu=1}^{16} \left( \sum_{a=1}^{4} \sum_{\mu=1}^{16} \text{sgn} \left( \frac{3}{2} - a \right) \frac{1}{2\sqrt{T_a}} \langle \xi_a | \hat{q}_{\mu,\nu} s_{\mu} | \zeta_a \rangle \right)^2 \Lambda_{\nu}.
\] (75)

For our example the concurrence is 0.963 ± 0.018.

Once we know the error in the concurrence, the errors in the tangle and the entanglement of formation can be found straightforwardly:

\[
\Delta T = 2C \Delta C, \quad \Delta E = \frac{C}{\sqrt{1 - C^2}} h' \left( \frac{1 + \sqrt{1 - C^2}}{2} \right) \Delta C,
\]

where \( h'(x) \) is the derivative of \( h(x) \). For our example the tangle is 0.928 ± 0.034 and the entanglement of formation is 0.947 ± 0.025.

6. Conclusions

In conclusion, we have presented a technique for reconstructing density matrices of qubit systems, including a full error analysis. We have extended the latter through to calculation of quantities of interest in quantum information, such as the entropy and concurrence. Without loss of generality, we have used the example of polarization qubits of entangled photons, but we stress that these techniques can be adapted to any physical realization of qubits.

Acknowledgements

The authors would like to thank Joe Altepeter, Mauro d’Ariano, Zdenek Hradil, Susana Huelga, Kurt Jacobs, Poul Jessen, Jian Li, James D. Malley, Michael Nielsen, Mike Raymer, Sze Tan, and Jaroslav Reháček for useful discussions and correspondence. This work was supported in part by the U.S. National Security Agency, and Advanced Research and Development Activity (ARDA), by the Los Alamos National Laboratory LDRD program, and by the Australian Research Council.
Appendix

A. The \( \hat{\Gamma} \)-Matrices

One possible set of \( \hat{\Gamma} \)-matrices are generators of \( SU(2) \otimes SU(2) \), normalized so that the conditions given in Eq. (18) are fulfilled. These matrices are

\[
\hat{\Gamma}_1 = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \hat{\Gamma}_2 = \frac{1}{2} \begin{pmatrix} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \hat{\Gamma}_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \\
\hat{\Gamma}_4 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad \hat{\Gamma}_5 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \quad \hat{\Gamma}_6 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \\ i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix}, \\
\hat{\Gamma}_7 = \frac{1}{2} \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \hat{\Gamma}_8 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \quad \hat{\Gamma}_9 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & i & 0 \end{pmatrix}, \\
\hat{\Gamma}_{10} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \hat{\Gamma}_{11} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \hat{\Gamma}_{12} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
\hat{\Gamma}_{13} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad \hat{\Gamma}_{14} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \hat{\Gamma}_{15} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
\hat{\Gamma}_{16} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.
\]

As noted in the text, this is only one possible choice for these matrices, and the final results are independent of the choice.

B. The \( \hat{M} \)-Matrices and Some of Their Properties

The \( \hat{M} \) matrices, defined by Eq. (25), are as follows:

\[
\hat{M}_1 = \frac{1}{2} \begin{pmatrix} 2 & -(1 - i) & -(1 + i) & 1 \\ -(1 + i) & 0 & i & 0 \\ -(1 - i) & -i & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \hat{M}_2 = \frac{1}{2} \begin{pmatrix} 0 & -(1 - i) & 0 & 1 \\ -(1 + i) & 2 & i & -(1 + i) \\ 0 & -i & 0 & 0 \\ 1 & -(1 - i) & 0 & 0 \end{pmatrix}.
\]
\[ M_3 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & i & -(1+i) \\ 0 & -i & 0 & -(1-i) \\ 1-(1-i) & -(1+i) & 2 \end{pmatrix}, \quad M_4 = \frac{1}{2} \begin{pmatrix} 0 & 0 & -(1+i) & 1 \\ 0 & 0 & i & 0 \\ -(1-i) & -(1+i) & 2 \\ 1 & 0 & -(1+i) & 0 \end{pmatrix}, \]

\[ M_5 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 2i & -(1+i) \\ 0 & 0 & 0 & (1-i) \\ -2i & (1+i) & 0 & 0 \\ -(1-i) & 0 & 0 & 2i \end{pmatrix}, \quad M_6 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -(1+i) \\ 0 & 0 & (1+i) & 2i \\ 0 & -(1-i) & 0 & 0 \\ 0 & 0 & -(1+i) & 0 \end{pmatrix}, \]

\[ M_7 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -(1+i) \\ 0 & 0 & -(1-i) & 2 \\ 0 & -(1+i) & 0 & 0 \\ -(1-i) & 0 & 0 & 0 \end{pmatrix}, \quad M_8 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -(1+i) \\ 0 & 0 & -(1-i) & 0 \\ 0 & -(1+i) & 0 & 0 \\ -(1-i) & 0 & 0 & 0 \end{pmatrix}, \]

\[ M_{11} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad M_{12} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -(1+i) \\ 0 & 0 & 0 & 0 \\ -(1-i) & 0 & 0 & 0 \end{pmatrix}, \]

\[ M_{13} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -(1+i) \\ 0 & 0 & -(1-i) & 0 \\ 0 & -(1+i) & 0 & 2 \\ -(1-i) & 0 & 2 & 0 \end{pmatrix}, \quad M_{14} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & -(1+i) \\ 0 & 0 & -(1-i) & 0 \\ 0 & (1+i) & 0 & 2i \\ -(1+i) & 0 & 2i & 0 \end{pmatrix}, \]

\[ M_{15} = \frac{1}{2} \begin{pmatrix} 0 & 0 & -(2i) & 0 \\ 0 & 0 & 0 & -(1-i) \\ 2i & 0 & -(1+i) & 0 \\ -(1+i) & 0 & 0 & 0 \end{pmatrix}, \quad M_{16} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \]

The form of these matrices is independent of the chosen set of matrices \( \{ \hat{\Gamma}_\nu \} \) used to convert the density matrix into a column vector. However, the \( \hat{M}_{\nu} \) matrices do depend on the set of tomographic states \( |\psi_\nu \rangle \).

There are some useful properties of these matrices which we will now derive. From Eq. (25), we have

\[ \langle \psi_\mu | \hat{M}_\nu | \psi_\mu \rangle = \sum_\lambda \langle \psi_\mu | \hat{\Gamma}_\lambda | \psi_\mu \rangle (B^{-1})_{\lambda,\nu}. \] (76)

From Eq. (22) we have \( \langle \psi_\mu | \hat{\Gamma}_\lambda | \psi_\mu \rangle = B_{\mu,\lambda} \); thus we obtain the result

\[ \langle \psi_\mu | \hat{M}_\nu | \psi_\mu \rangle = \delta_{\mu,\nu}. \] (77)

If we denote the basis set for the four-dimensional Hilbert space by \( \{|i\rangle (i = 1, 2, 3, 4)\} \), then Eq. (24) can be written as follows:

\[ \langle i | \hat{\rho} | j \rangle = \sum_{k,l} \sum_\nu \langle i | \hat{M}_\nu | j \rangle \langle \psi_\nu | k \rangle \langle l | \psi_\nu \rangle \langle k | \hat{\rho} | l \rangle. \] (78)
Since Eq. (78) is valid for arbitrary states $\hat{\rho}$, we obtain the following relationship:

$$
\sum_{\nu} \langle i | \hat{M}_{\nu} | j \rangle \langle \psi_{\nu} | k \rangle \langle l | \psi_{\nu} \rangle = \delta_{ik} \delta_{jl}.
$$  

(79)

Contracting Eq. (79) over the indices $(i, j)$ we obtain:

$$
\sum_{\nu} \text{Tr} \left\{ \hat{M}_{\nu} \right\} |\psi_{\nu} \rangle \langle \psi_{\nu}| = \hat{I},
$$  

(80)

where $\hat{I}$ is the identity operator for our four dimensional Hilbert space.

A second relationship can be obtained by contracting Eq. (79), viz.,

$$
\sum_{\nu} \langle i | \hat{M}_{\nu} | j \rangle = \delta_{ij},
$$  

(81)

or, in operator notation,

$$
\sum_{\nu} \hat{M}_{\nu} = \hat{I}.
$$  

(82)

C. Perturbation Theory for Non-Hermitian Matrices

Whereas perturbation theory for Hermitian matrices is covered in most quantum mechanics textbooks, the case of non-Hermitian matrices is less familiar, and so we will present it here. The problem is as follows. Given the eigenspectrum of a matrix $\hat{R}_0$ [32], i.e.,

$$
\langle \xi_{\alpha} | \hat{R}_0 = r_{\alpha} \langle \xi_{\alpha} |, \quad \hat{R}_0 |\zeta_{\alpha} \rangle = r_{\alpha} |\zeta_{\alpha} \rangle,
$$  

we wish to find expressions for the eigenvalues $r'_{\alpha}$ and eigenstates $\langle \xi'_{\alpha} |$ and $|\zeta'_{\alpha} \rangle$ of the perturbed matrix $\hat{R}' = \hat{R}_0 + \delta \hat{R}$.

We start with the standard assumption of perturbation theory, i.e., that the perturbed quantities $r'_{\alpha}$, $\langle \xi'_{\alpha} |$, and $|\zeta'_{\alpha} \rangle$ can be expressed as power series of some parameter $\lambda$:

$$
r'_{\alpha} = r^{(0)}_{\alpha} + \lambda r^{(1)}_{\alpha} + \lambda^2 r^{(2)}_{\alpha} + \cdots,
$$  

(84)

$$
|\zeta'_{\alpha} \rangle = |\zeta^{(0)}_{\alpha} \rangle + \lambda |\zeta^{(1)}_{\alpha} \rangle + \lambda^2 |\zeta^{(2)}_{\alpha} \rangle + \cdots,
$$  

(85)

$$
\langle \xi'_{\alpha} | = \langle \xi^{(0)}_{\alpha} | + \lambda \langle \xi^{(1)}_{\alpha} | + \lambda^2 \langle \xi^{(2)}_{\alpha} | + \cdots.
$$  

(86)
Writing \( \hat{R}' = \hat{R}_0 + \lambda \delta \hat{R} \), and comparing terms of equal powers of \( \lambda \) in the eigen-equations, one obtains the following formulas:

\[
\hat{R}_0 |\zeta_a^{(0)}\rangle = r_a^{(0)} |\zeta_a^{(0)}\rangle, \quad (87)
\]

\[
\langle \zeta_a^{(0)} | \hat{R}_0 = r_a^{(0)} \langle \zeta_a^{(0)} |. \quad (88)
\]

\[
(\hat{R}_0 - r_a^{(0)} \hat{I}) |\zeta_a^{(1)}\rangle = - \left( \delta \hat{R} - r_a^{(1)} \right) |\zeta_a^{(0)}\rangle, \quad (89)
\]

\[
\langle \xi^{(1)}_a | (\hat{R}_0 - r_a^{(0)} \hat{I}) = - \langle \xi^{(0)}_a | \left( \delta \hat{R} - r_a^{(1)} \right). \quad (90)
\]

Equations (87) and (88) imply that, as might be expected,

\[
|\zeta_a^{(0)}\rangle = |\zeta_a\rangle, \quad \langle \xi^{(0)}_a | = \langle \xi_a |, \quad r_a^{(0)} = r_a.
\]

Taking the inner product of Eq. (89) with \( \langle \xi_a | \), and using the bi-orthogonal property Eq. (83), we obtain

\[
r_a^{(1)} = \langle \xi_a | \delta \hat{R} |\zeta_a\rangle. \quad (91)
\]

This implies that

\[
\delta r_a \equiv r_a' - r_a \approx \langle \xi_a | \delta \hat{R} |\zeta_a\rangle. \quad (92)
\]

Thus, dividing both sides by some differential increment \( \delta x \) and taking the limit \( \delta x \to 0 \), we obtain

\[
\frac{\partial r_a}{\partial x} = \langle \xi_a | \frac{\partial \hat{R}}{\partial x} |\zeta_a\rangle. \quad (93)
\]

Using the completeness property of the eigenstates, \( \sum_b |\zeta_b\rangle \langle \xi_b | = \hat{I} \), and the identity \( \hat{R}_0 = \sum_b r_b |\zeta_b\rangle \langle \xi_b | \) we obtain the following formula

\[
(\hat{R}_0 - r_a \hat{I})^{-1} = \sum_{b \neq a} \frac{1}{r_b - r_a} |\zeta_b\rangle \langle \xi_b |. \quad (94)
\]

Applying this to Eq. (89) we obtain

\[
|\delta \zeta_a^{(1)}\rangle \equiv |\zeta_a'^{\prime}\rangle - |\zeta_a\rangle \approx - \sum_{b \neq a} \left( \frac{\langle \xi_b | \delta \hat{R} |\zeta_a\rangle}{r_b - r_a} \right) |\zeta_b\rangle.
\]

Similarly, Eqs. (90) and (94) imply

\[
\langle \delta \xi_a | \equiv \langle \delta \xi_a'^{\prime} | - \langle \delta \xi_a | \approx - \sum_{b \neq a} \left( \frac{\langle \xi_a | \delta \hat{R} |\zeta_b\rangle}{r_b - r_a} \right) \langle \xi_b |. \quad (95)
\]
References


INDEX

*-subalgebra, 120
$C^k$-state manifold, 151
$C^k$-map, 151
G–locally efficient, 237
G-efficient, 237
m-semiclassical, 373
n-i.i.d. extended family, 173
$\delta$-typical, 87, 88
G-efficient, 237
(first-order) asymptotically efficient, 128

acceptance region, 424
AF $C^*$-algebra, 46, 56, 61
anti-normal ordered, 503
anti-symmetric logarithmic derivative, 222, 249
antiunitary, 347
approximate cloning, 458
asymptotic efficiency, 129
asymptotic strong orthogonality, 23, 83
asymptotic weak orthogonality, 23, 78
asymptotically exact, 461
asymptotically unbiased, 164
attainable Cramér-Rao type bound, 92, 317
auxiliary qubits, 379

base space, 311
Bayesian approach, 9
Bayesian reconstruction, 474
Black Cow factor, 380
Bloch vector, 379, 380

character, 464
coherent, 292, 333, 335
coherent state, 141
collective measurement, 94
commutation operator, 289
commutative, 68, 123
commuting theorem, 323
completely ergodic, 46, 48
completely positive map, 381
completeness, 141
composite systems, 408
concatenation, 380, 383
conditional expectation, 122
consistent, 128
consistent estimator, 308, 413
converse part, 21
covariant, 369, 370
CR bound, 104
Cramér-Rao inequality, 96
Cramér-Rao type bound, 320
density operator, 101, 116, 137, 230, 234
deviation, 153
direct approach, 306, 321
direct part, 21
direct sum representation, 71
disjoint random combination, 429
distinguishable, 23
duality, 305
e-tangent vector, 234
effective quantum efficiency, 506
efficient, 96, 126, 227, 237, 253
efficient estimator, 118
ergodic, 47
estimating the spectrum, 458
estimation vector, 322
estimator, 103, 163, 224, 230, 309, 310, 411
exponential connection, 129
exponential error estimate, 463
exponential representation, 234
faithful, 151, 441
faithful model, 318
family, 91
family of probability distributions, 8, 91
fidelity, 379, 381, 382
first error probability, 46
first order asymptotic theory, 162
Fisher information, 91, 121
Fisher information matrix, 92, 308
Fubini-Study distance, 368
Gärtner-Ellis theorem, 464
Gaussian displacement noise, 505
Gaussian state, 274
generalized measurement, 134, 230
generalized quantum measurement, 115
generalized spin coherent model, 333
GNS representation, 53, 62
half-line quantum Gaussian state
family, 431
highest weight, 463
Homodyne, 472
homodyning the observable, 500
horizontal component, 312
horizontal lift, 312, 313, 344
horizontal subspace, 311
horizontal vector, 312
hyperfinite, 57
hyperfinite von Neumann algebra, 46, 61
indirect approach, 321
infinite $C^*$-tensor product, 45
information divergence, 43
information geometry, 305
informationally exclusive, 339
informationally independent, 338, 340
invariance approach, 9
invariant, 71
irreducible, 71
KMB Fisher information, 396
Kolmogorov-Sinai entropy, 53, 62
large deviation, 460
large deviation method, 463
large deviation theory, 395
likelihood ratio test, 19
linear tomography, 474
local unbiasedness, 139
locally coherent, 292
locally efficient, 237
locally quasi-classical, 318
locally unbiased, 152, 225, 231, 310, 322
locally unbiased at, 104
locally unbiased estimator, 307
locally unbiased operators, 231
LOCC, 23
logarithmic derivative, 91
lower bound of Cramér-Rao type, 104
m-tangent vector, 234
martingale convergence of relative
entropy, 56
matrix representation, 311
maximum likelihood estimator
(MLE), 98
maximum likelihood tomography, 474
mean ergodic theorem, 46, 53, 62
mean relative entropy, 45, 47
mean square error (MSE), 91
measurement, 102, 115, 134, 137, 309, 407
minimally uncertain state, 141
minimax, 9
minimum variance unbiased
estimation, 9
mixed, 101
mixture connection, 129
mixture representation, 234
model, 162, 307, 309
monotonicity, 20
monotonicity of relative entropy, 44
most informative CR bound, 107
most informative Cramér-Rao bound, 92
motion reversal operator, 348
MSE consistent, 164
MSE consistent estimator, 173
multiplicity, 464
mutually dual, 130

Neyman Pearson Lemma, 19, 24
non-commutative Fourier transform, 272
non-quantum-correlational Cramér-Rao type bound, 92, 171
normal ordered, 502
normalized $\rho$θ-symplectic basis, 294

observable, 103, 408
observation, 44
one-parameter quantum Gaussian state family, 431
operator-concavity, 50, 61
operator-valued relative entropy, 49
optimal, 466
optimal quantum cloning, 385
orthogonal resolution of identity, 115
outcome of the measurement, 137
over-completeness, 141

parallel, 319
POM, 13
position shifted model, 346, 348
positive operator-valued measure, 13, 44, 163, 368, 461
POVM, 13, 368, 382
pre-inner product, 221, 247
principal fiber bundle, 311
probability operator measure, 13, 136
probability operator-valued measure, 102
product state, 45
projection-valued, 102
projection-valued measure, 44
pure, 101, 221, 230, 247
pure states, 418
quantum birthday problems, 77
quantum channel, 487
quantum characteristic function, 273
quantum cloning, 379, 380
quantum Cramér-Rao inequality, 114
quantum Cramér-Rao type bound, 168
quantum exponential family, 119, 127
quantum hypothesis testing, 45
quantum i.i.d., 10, 14
quantum information geometry, 305
quantum Neyman Pearson Lemma, 19
quantum relative entropy, 20
quantum state family, 92
quantum statistical model, 103
Quantum Stein’s Lemma, 21
quantum-correlational Cramér-Rao type bound, 173
quasi-classical, 94, 121, 318
quasi-classical in the wider sense, 345
quasi-quantum Cramér-Rao type bound, 164
random combination, 429
rate function, 463
recursive estimator, 171
recursive MSE consistent estimator, 171
relative entropy, 20, 43, 130
relative phase factor, 313
representation, 71
right logarithmic derivative, 93, 104, 117, 232, 295
RLD, 93
RLD Fisher information matrix, 93, 172
scaled cumulant generating function, 463
second error probability, 46
Shannon-McMillan-Breiman theorem, 46, 53, 57, 63
shrinking factor, 380, 381, 383, 384
sieved maximum likelihood estimation, 473
simple, 102, 230
simple hypothesis testing, 19
simple measurement, 134, 407
SLD CR bound, 314
SLD CR inequality, 314
SLD Fisher information, 92, 396
SLD Fisher information matrix, 223, 250, 314
SLD tangent space, 292
small boundary, 463
spectral decomposition, 115
spin coherent model, 338
state, 44, 101, 115, 116, 137, 406
state change caused by the measurement, 407
stationary, 47
statistical operator, 116, 137
statistics of the measurement result, 407
Stein’s Lemma, 20, 24
Stokes parameters, 512
strong consistency, 397
strong orthogonality, 76
strong orthogonality capacity, 76
strongly clustering, 47
strongly consistent, 434
strongly consistent at \( \theta \), 434
strongly mixing, 47
structure group, 311
sufficiency, 122
sufficient, 122, 123
superadditivity of relative entropy, 47
symmetric logarithmic derivative, 92, 104, 117, 222, 231, 249, 291, 401
symmetric subspace, 379, 384
symmetrized logarithmic derivative, 313
tensored representation, 71
test, 410
time reversal operator, 348
tomography, 472
total space model, 338
trace, 44
Uhlmann fidelity, 384
Uhlmann’s relative entropy, 55
Umegaki’s relative entropy, 44
unbiased, 95, 103, 138, 225, 231, 310
unbiased estimator, 116, 307
unbiasedness, 411, 422
uncertainty principle, 134
universal \( N \rightarrow M \) cloner, 385
universal \( N \rightarrow M \) quantum cloning machine, 380
universal and symmetric quantum cloning machine, 379
universal cloning machine, 381, 384
Varadhan’s theorem, 464
vertical component, 312
vertical subspace, 311
vertical vector, 312
weak consistency, 423, 426
weak orthogonality, 76
weak orthogonality capacity, 76
weakly consistent, 396, 431
weakly mixing, 47
weight field, 126
weight matrix, 317
weights, 464
Weyl representation, 272
Wigner function, 472
Young frame, 459, 460